CUK-H03257-36-023942

76

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

WEI-LIANG CHOW
THE JOHNS HOPKINS UNIVERSITY

PHILIP HARTMAN
THE JOHNS HOPKINS UNIVERSITY

A. M. GLEASON
HARVARD UNIVERSITY

GILBERT A. HUNT
PRINCETON UNIVERSITY

WITH THE COÖPERATION OF

L. V. AHLFORS
A. BOREL
H. CARTAN
S. S. CHERN
C. CHEVALLEY
K. L. CHUNG
J. A. DIEUDONNÉ
K. IWASAWA
K. KODAIRA
F. I. MAUTNER
J. MILNOR
A. WEIL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

510.5
.002
84

CALCUTTA UNIVERSITY
CENTRAL LIBRARY

VOLUME LXXXIV

1962

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

IV

239342

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

Artin, Michael. Some numerical criteria for contractability of curves on algebraic surfaces, ... 485
Auslander, Maurice. A remark on a paper of M. Hironaka, . . . 8
Auslander, Maurice. On the purity of the branch locus, . . . 116
Białynicki-Birula, A. On Galois theory of fields with operators, . 89
Bing, R. H. Each disk in E^3 contains a tame arc, ... 583
Bing, R. H. Each disk in E^3 is pierced by a tame arc, ... 591
Birch, B. J., D. J. Lewis, and T. G. Murphy. Simultaneous quadratic forms, ... 110
Branges, Louis de. Perturbations of self-adjoint transformations, . 543
Briney, R. E. Intersection theory on quotients of algebraic varieties, 217
Browder, Felix E. Approximation by solutions of partial differential equations, ... 134
Browder, Felix E. Analyticity and partial differential equations, I, 666
Cohn, Harvey. Cusp forms arising from Hilbert's modular functions for the field of $3^{\frac{1}{3}}$, ... 283
Davenport, H. and D. J. Lewis. Exponential sums in many variables, 649
Dyer, Eldon and R. K. Lashof. Homology of iterated loop spaces, . 35
Edwards, C. H., Jr. On the union of two solid tori, ... 11
Ehrenpreis, Leon. Solutions of some problems of division. Part V. Hyperbolic operators, ... 324
Falb, Peter L. On the behaviour of the geometric genus under ground field extension, ... 421
Feit, Walter. A characterization of the simple groups $SL(2,2^a)$: correction, ... 201
Gilbert, R. P. Some properties of generalized axially symmetric potentials, ... 475
Goldberg, Samuel I. and Shoshichi Kobayashi. The conformal transformation group of a compact Riemannian manifold, . 170
Gugenheim, V. K. A. M. On extensions of algebras, co-algebras and Hopf algebras, I, ... 349
Hartshorne, Robin. Complete intersections and connectedness, . 497
Heller, Alex, and K. A. Rowe. On the category of sheaves, . . . 205
Herz, C. S. On the number of lattice points in a convex set, . . . 126
Hillman, Abraham P. and David G. Mead. On the Ritt polygon process, ... 629
Igusa, Jun-ichi. On Siegel modular forms of genus two, . . . 175
Jacobowitz, Ronald. Hermitian forms over local fields, ... 441

PAGE

Jenkins, James A. Remarks on the paper "Generalization of a theorem of Mandelbrojt," 383

Knopp, Marvin Isadore. On abelian integrals of the second kind and modular functions, 615

Kobayashi, Shoshichi. See Samuel I. Goldberg and Shoshichi Kobayashi, page 170.

Kwun, Kyung Whan and Frank Raymond. Factors of cubes, 433

Lashof, R. K. See Eldon Dyer and R. K. Lashof, page 35.

Lewis, D. J. See B. J. Birch, D. J. Lewis, and T. G. Murphy, page 110.

Lewis, D. J. See H. Davenport and D. J. Lewis, page 649.

Loomis, L. H. Unique direct decompositions on convex sets, 509

McQuillan, Donald L. A generalization of a theorem of Hecke, 306

Mead, David G. See Abraham P. Hillman and David G. Mead, page 629.

Milnor, J. A unique decomposition theorem for 3-manifolds, 1

Mostow, G. D. Covariant fiberings of Klein spaces, II, 466

Muckenhoupt, Benjamin. Some results on analytic iteration and conjugacy, 161

Mumford, David. Further pathologies in algebraic geometry, 642

Murphy, T. G. See B. J. Birch, D. J. Lewis, and T. G. Murphy, page 110.

Raymond, Frank. See Kyung Whan Kwun and Frank Raymond, page 433.

Robertson, M. M. Partitions of large multipartites, 16

Rodrigues, A. M. The first and second fundamental theorems of Lie for Lie pseudo groups, 265

Rowe, K. A. See Alex Heller and K. A. Rowe, page 205.

Sandler, Reuben. Autotopism groups of some finite non-associative algebras, 239

Schlesinger, Ernest C. Solvable functions for the Dirichlet problem, 317

Smale, S. On the structure of manifolds, 387

Treves, François. Fundamental solution of linear partial differential equations with constant coefficients depending on parameters, 561

Tucker, Patricia A. On the reduction of induced representations of finite groups, 400

Venkatesan, P. S. On a class of inverse semigroups, 578

Wolf, Joseph A. Discrete groups, symmetric spaces, and global holonomy, 527

Wonenburger, María J. The automorphisms of the group of similitudes and some related groups, 600

Wonenburger, María J. The automorphism of $PO_8^+(Q)$ and $PS_8^+(Q)$, 635

CENTRAL LIBRARY UNIVERSITY

A UNIQUE DECOMPOSITION THEOREM FOR 3-MANIFOLDS.*

By J. Milnor.

In 1928 Kneser proved that every closed 3-dimensional manifold can be built up in a more or less unique way out of "irreducible" 3-manifolds. (See [3, pg. 256].) This result has gained new interest in view of the sphere theorem of Papakyriakopoulos and Whitehead [7 and 10] which yields a partial characterization of the irreducible manifolds.

All manifolds considered are to be connected, oriented, triangulated, and without boundary. All imbeddings and homeomorphisms are to be piecewise linear. Two manifolds M and M' will be called *isomorphic* ($M \approx M'$) if there exists a piecewise linear, orientation preserving homeomorphism between them.

The *connected sum* $M \# M'$ of two 3-manifolds is obtained by removing the interior of a 3-cell from each, and then matching the resulting boundaries, using an orientation reversing homeomorphism. (Compare Seifert [9], Milnor [6].) It follows from the homogeneity theorem of Newman and Gugenheim [2, Theorem 3] that this operation is well defined up to isomorphism. It is clearly associative and commutative up to isomorphism. The sphere S^3 serves as identity element.

Closely related is the operation of "adding a handle" to a connected 3-manifold M. Choosing two disjoint 3-cells in M, one removes their interiors and matches the resulting boundaries under an orientation reversing homeomorphism. Again it follows from the Newman-Gugenheim theorem that this operation is well defined. If one adds a handle to S^3 the result is isomorphic to $S^1 \times S^2$. Hence if one adds a handle to any manifold $M \approx M \# S^3$ the result is isomorphic to $M \# (S^1 \times S^2)$.

It will be convenient to call a manifold *non-trivial* if it is not isomorphic to S^3.

Definition. A non-trivial 3-manifold P is *prime* if there is no decomposition $P \approx M_1 \# M_2$ with M_1 and M_2 non-trivial.

Theorem 1. *Every non-trivial compact 3-manifold M is isomorphic to a sum $P_1 \# \cdots \# P_k$ of prime manifolds. The summands P_i are uniquely determined up to order and isomorphism.*

---

* Received Feburary 7, 1961.

For the proof that M can be expressed as a sum of primes, I will refer to Kneser [3, pp. 252-255].

(If M itself is not prime then one has an expression $M \approx M_1 \# M_2$ with M_1 and M_2 non-trivial. If either M_1 or M_2 is not prime then one can express M as a three-fold sum, and so on. The difficulty is to show that this process must terminate after a finite number of steps. This Kneser proves by an ingenious argument. If one assumes the Poincaré hypothesis then there is a much easier proof. Define $\rho(M)$ as the smallest number of generators for the fundamental group of M. It follows from the Gruško-Neumann theorem [4, pg. 58] that $\rho(M_1 \# M_2) = \rho(M_1) + \rho(M_2)$. Hence if $M \approx M_1 \# \cdots \# M_k$ with $k > \rho(M)$ then some M_i must satisfy $\rho(M_i) = 0$, and hence must be isomorphic to S^3.)

For the rest of the proof we can drop the hypothesis that M is compact.

Kneser calls a manifold M *irreducible* if every 2-sphere in M bounds a 3-cell. For example the theorem of Alexander [1] asserts that S^3 is irreducible. On the other hand the product $S^1 \times S^2$ contains a 2-sphere $p \times S^2$ which does not bound any cell

LEMMA 1. *With the exception of manifolds isomorphic to S^3 or $S^1 \times S^2$, a manifold is prime if and only if it is irreducible.*

Proof. If $M = M_1 \# M_2$ with M_1 and M_2 non-trivial then the 2-sphere which separates the two summands in M cannot bound a cell. Conversely let $\Sigma \subset M$ be a 2-sphere which does not bound a cell. If Σ separates M, then clearly M can be expressed as a non-trivial sum. If Σ does not separate M, then cutting along Σ, and pasting in two 3-cells to eliminate the resulting boundaries we obtain a connected manifold M'. The original manifold can be obtained from M' by adding a handle; hence $M \approx M' \# (S^1 \times S^2)$. Therefore either M can be expressed non-trivially as a sum, or else $M \approx S^1 \times S^2$. This completes the proof.

LEMMA 2. *$S^1 \times S^2$ is prime.* It is sufficient to prove that every bounding 2-sphere in $S^1 \times S^2$ bounds a cell. Lifting the 2-sphere to the universal covering space, which can be imbedded in S^3, this follows from Alexander's theorem.

LEMMA 3. *Suppose that $M = M_1 \# M_2$ is isomorphic to a sum of primes $P_1 \# \cdots \# P_k$. Then the P_i can be renumbered so that*

$$M_1 \approx P_1 \# \cdots \# P_r \text{ and } M_2 \approx P_{r+1} \# \cdots \# P_k.$$

It is clear that Theorem 1 will follow from this lemma.

Proof. Consider manifolds bounded by finitely many 2-spheres. It will be convenient to call two such manifolds *equivalent* if one can be obtained from the other either by removing the interiors of finitely many disjoint interior 3-cells, or by the converse operation of filling in 3-cells.

Suppose that the first s manifolds $P_1, \cdots, P_s$ are irreducible, and that the remaining manifolds $P_{s+1}, \cdots, P_k$ are isomorphic to $S^1 \times S^2$. Let T denote the 2-sphere which separates the two halves of $M = M_1 \# M_2$.

Case 1. $s > 0$. Consider systems $\{\Sigma_1, \cdots, \Sigma_{k-1}\}$ of disjoint 2-spheres in M which satisfy:

Condition 1) the intersections $S \# T \cap \Sigma_n$ consist of disjoint simple closed curves; and

Condition 2) cutting M along the Σ_n we obtain a manifold-with-boundary having s components $B_1, \cdots, B_s$ which are equivalent to $P_1, \cdots, P_s$ respectively.

First it is necessary to show that such systems exist. Thinking of M as $P_1 \# \cdots \# P_s$ with $k - s$ handles attached, let $\Sigma_1, \cdots, \Sigma_{s-1}$ separate the s summands and let $\Sigma_s, \cdots, \Sigma_{k-1}$ cut the $k - s$ handles. By a "general position" argument we may assume that Condition 1 is satisfied. Thus $\{\Sigma_1, \cdots, \Sigma_{k-1}\}$ satisfies both conditions.

Now suppose that $\Sigma_1 \cup \cdots \cup \Sigma_{k-1}$ does actually intersect T. We will construct a new system $\{\Sigma_1, \cdots, \Sigma_h', \cdots, \Sigma_{k-1}\}$ with fewer intersection curves.

Among the intersection curves $C \subset T \cap \Sigma_h$ choose one which bounds a 2-cell $E \subset T$ containing no other intersection curves. Then E is contained in one of the manifolds B_i. Let B_j denote the manifold which lies on the other side of Σ_h from B_i. (Caution: it may happen that $i = j$. Compare Figures 1, 2.)

Since B_i is equivalent to the irreducible manifold P_i, the cell E spanning one of its boundary 2-spheres must cut B_i into two parts, B_i' and B_i'', one of which is equivalent to S^3. Let E' and E'' denote the corresponding 2-cells in Σ_h, bounded by C.

Case 1.1. $i \neq j$. Suppose that B_i'' is the part of B_i which is equivalent to S^3. Consider the 2-sphere Σ_h' formed from $E \cup E'$ by deforming slightly into B_i'. Then Σ_h' has fewer intersection curves with T than Σ_h. If Σ_h is replaced by Σ_h' then the effect on the complementary components is to subtract B_i'' from B_i and to add it to B_j. Since B_i'' is equivalent to S^3, this does not change the equivalence classes of B_i and B_j.

Case 1.2. $i = j$. Then one of the two parts of B_i, say B_i', must contain the "other side" of Σ_h. (Compare Figure 2.) Again let Σ_h' be obtained by deforming $E \cup E'$ slightly into B_i'. If Σ_h is replaced by Σ_h' then the effect on B_i is to subtract B_i'' from one part of B_i, but then to add it on again along a different boundary component. This procedure clearly does not alter the equivalence of class of B_i. Hence $\{\Sigma_1, \cdots, \Sigma_h', \cdots, \Sigma_{k-1}\}$ still satisfies Conditions 1 and 2.

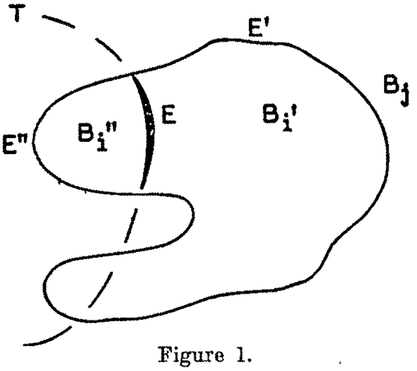

Figure 1.

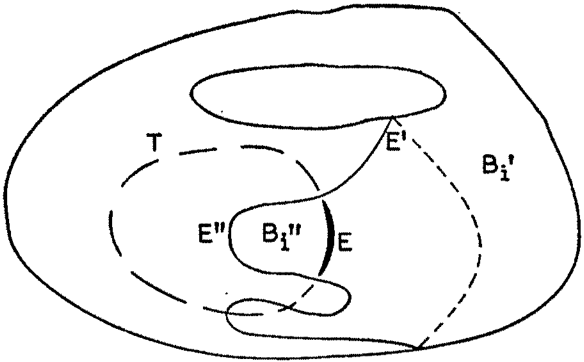

Figure 2.

Thus we can eliminate intersection curves one by one until we obtain a system $\{\Sigma_1, \cdots, \Sigma_{k-1}\}$ which is completely disjoint from T.

Suppose that T lies within B_i, and cuts it into B_i' and B_i'', where say B_i'' is equivalent to S^3. Suppose further that $B_1, \cdots, B_{i-1}, B_i'$ lie on the M_1 side of T and that $B_i'', B_{i+1}, \cdots, B_s$ lie on the M_2 side. Then clearly M_1 is isomorphic to $P_1 \# \cdots \# P_i$ with a certain number of handles attached, and M_2 is isomorphic to $P_{i+1} \# \cdots \# P_s$ with the remaining number of handles attached. This completes the proof in Case 1.

Case 2. All of the primes $P_1, \cdots, P_k$ are isomorphic to $S^1 \times S^2$. Then the proof is similar except that one uses k disjoint spheres (instead of $k-1$) and Condition 2 must be replaced by the following.

Condition 2′. Cutting M along the spheres $\Sigma_1, \cdots, \Sigma_k$ we obtain a connected manifold-with-boundary B which is equivalent to S^3.

Otherwise there is no change in the argument. This completes the proof of Lemma 3, and of Theorem 1.

Thus the classification problem for closed 3-manifolds is reduced to the problem of classifying prime (or irreducible) manifolds. As a consequence of the sphere theorem we have.

THEOREM 2. *For every irreducible manifold M the homotopy group $\pi_2(M)$ is zero. If the Poincaré hypothesis is true, then conversely every manifold with $\pi_2 = 0$ is irreducible.*

Proof. If $\pi_2(M) \neq 0$ then the sphere theorem asserts that there is a 2-sphere $\Sigma \subset M$ representing a non-zero element of $\pi_2(M)$. Certainly Σ cannot bound a cell. Conversely if $\pi_2(M) = 0$ then any $\Sigma \subset M$ must cut M into two parts, say B_1 and B_2. At least one of these parts must be compact and simply connected. For otherwise, lifting Σ to the universal covering space $\tilde{M}$, it would cut $\tilde{M}$ into two non-compact parts, and hence would not bound. But the Poincaré hypothesis implies that a compact simply connected manifold bounded by a sphere is a cell. This completes the proof.

Manifolds with $\pi_2 = 0$ can be divided into two classes, according as the universal covering space $\tilde{M}$ is compact or not. In the first case M itself must be compact and π_1 must be finite. The universal covering space has the homotopy type of S^3. The problem of classifying such manifolds seems difficult (compare [5]).

If the universal covering space $\tilde{M}$ is not compact then $\tilde{M}$ is contractible, and all of the higher homotopy groups $\pi_i(M)$, $i > 1$, are zero. Using the Smith theory it follows that $\pi_1(M)$ has no elements of finite order. (See Papakyriakoupoulos [7].) Such a manifold M may be either compact (for example a circle bundle over a surface of genus $\geqq 1$) or non-compact (for example the complement of a knot in S^3). The problem of classifying such manifolds seems hopelessly difficult.

Another characterization of prime manifolds is due to Kneser [3] and Whitehead [11]: Modulo the Poincaré hypothesis, a closed manifold is prime if and only if its fundamental group can not be expressed as a free product in a non-trivial way.

In conclusion here are four possible generalizations of the above theory.

1. Consider 3-manifolds with boundary, and define the sum operation using interior 3-cells. Then Theorems 1 and 2 go through without essential change.

2. Consider 3-manifolds with connected boundary and form the sum of two such manifolds by pasting together 2-cells on the boundary. Thus the closed 3-cell serves as identity element. *Problem*: Do analogues of Theorems 1 and 2 hold in this case?

3. Consider non-orientable 3-manifolds. Then Theorem 1 becomes false (although Kneser's formulation of this theorem remains true). As an example let N denote the twisted S^2 bundle over S^1. This is the non-orientable analogue of a handle. Then $N \# N \approx N \# (S^1 \times S^2)$ although $N \not\approx S^1 \times S^2$. Similarly Theorem 2 becomes false: The product of a projective plane P^2 and a circle is irreducible, although $\pi_2(P^2 \times S^1) \neq 0$.

4. Consider closed oriented 4-manifolds. Then again Theorem 1 becomes false. Let P denote the complex projective plane with the usual orientation and P' the same manifold with the opposite orientation. Then Hirzebruch has proved $P \# P' \# P'$ is diffeomorphic to $(S^2 \times S^2) \# P'$. But $P \# P'$ does not even have the same homotopy type as $S^2 \times S^2$.

To prove Hirzebruch's assertion consider the graph G of the algebraic correspondence $P \leftrightarrow S^2 \times S^2$ defined by

$$(z_0 : z_1 : z_2) \leftrightarrow (z_0 : z_1), (z_0 : z_2).$$

The induced correspondence $P \rightarrow G$ is biregular except that the two points $(0:0:1)$ and $(0:1:0)$ are "blown up" into 2-spheres. It follows that

$$G \approx P \# P' \# P'.$$

Similarly the correspondence $S^2 \times S^2 \rightarrow G$ is biregular except that the point $(0:1), (0:1)$ is blown up. Hence $G \approx (S^2 \times S^2) \# P'$; which completes the proof.

PRINCETON UNIVERSITY.

REFERENCES.

[1] J. W. Alexander, "On the subdivision of 3-space by a polyhedron," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 10 (1924), pp. 6-8.

[2] V. K. A. M. Gugenheim, "Piecewise linear isotopy and embedding of elements and spheres (I)," *Proceedings of the London Mathematical Society*, vol. 3 (1953), pp. 29-53.

[3] H. Kneser, "Geschlossen Flächen in dreidimensionalen Mannigfaltigkeiten," *Jahresbericht der Deutschen Mathematiker Vereinigung*, vol. 38 (1929), pp. 248-260.

[4] A. G. Kurosh, *The theory of groups*, vol. II, Chelsea (New York), 1956.

[5] J. Milnor, "Groups which act on S^n without fixed points," *American Journal of Mathematics*, vol. 79 (1957), pp. 623-630.

[6] ———, "Sommes de variétés différentiables et structures différentiables des sphères," *Bulletin de la Société Mathématique de France*, vol. 87 (1959), pp. 439-444.

[7] C. Papakyriakopoulos, "On Dehn's lemma and the asphericity of knots," *Annals of Mathematics*, vol. 66 (1957), pp. 1-26.

[8] ———, "Some problems on 3-dimensional manifolds," *Bulletin of the American Mathematical Society*, vol. 64 (1958), pp. 317-335.

[9] H. Seifert, "Konstruktion dreidimensionaler gescholossener Räume," *Berichte über die Verhandlung der Sächsischen Akademie der Wissenschaften zu Leipzig*, vol. 83 (1931), pp. 26-66.

[10] J. H. C. Whitehead, "On 2-spheres in 3-manifolds," *Bulletin of the American Mathematical Society*, vol. 64 (1958), pp. 161-166.

[11] ———, "On finite cocycles and the sphere theorem," *Colloquium Mathematicum*, vol. 6 (1959), pp. 271-281.

A REMARK ON A PAPER OF M. HIRONAKA.*

By Maurice Auslander.

In [4] H. Hironaka quoted and used, but did not prove, a theorem I had communicated to him which he called Theorem 0 in his paper. In this note I present a proof of this theorem.

We begin with a brief summary of some pertinent homological facts. The reader is referred to [1] for a more complete exposition.

Let R be a local ring and M a finitely generated non-zero R-module. A sequence of non-units $x_1, \cdots, x_t$ in R is said to be an M-sequence if x_i is not a zero divisor in $M/(x_1, \cdots, x_{i-1})M$ for all $i = 1, \cdots, t-1$. It is well known that if $x_1, \cdots, x_t$ is an M-sequence, then $t \leqq \dim(M) = \dim(R/\mathfrak{a}(M))$ where $\mathfrak{a}(M)$ is the annihilator of M. Thus it makes sense to talk about maximal M-sequences. It can be shown that all maximal M-sequences have the same length and this length is called the codimension of M (notation: $\operatorname{codim}_R M$). For an R-module M we define $d(M) = \dim(M) - \operatorname{codim} M$. It follows from our previous remark that $d(M) \geqq 0$.

Now if R is a regular local ring, then it is well known that the $\operatorname{hd} M + \operatorname{codim} M = \dim R$ for any finitely generated non-zero R-module M. Therefore $d(M) = \dim(M) - (\dim R - \operatorname{hd} M) = \operatorname{hd} M - \operatorname{rank} M$ where $\operatorname{rank} M = \operatorname{rank}(\mathfrak{a}(M))$ by definition.

Proposition 1. *Let R be a regular local ring, M a non-zero finitely generated R-module and t an integer greater than or equal to 0. Then $d(M) \geqq t$ if and only if there exist integers $i \geqq 0$ and $h \geqq t$ such that* $\operatorname{Ext}^i_R(M, R) \neq 0$ *and* $\operatorname{Ext}^{i+h}_R(M, R) \neq 0$.

Proof. We first recall a result of D. Rees (see [3]).

If S is a noetherian ring, M a finitely generated S-module and i the smallest integer such that $\operatorname{Ext}^i_S(M, S) \neq 0$, then i is the length of the largest S-sequence contained in the $\mathfrak{a}(M)$. If S is a regular local ring, then we have that $i = \operatorname{rank}(\mathfrak{a}(M))$ (see [1, Proposition 2.7]).

Now we return to the proof of the proposition. Suppose that $d(M) \geqq t$. Then we have that the $\operatorname{hd} M \geqq \operatorname{rank}(M) + t$. By the previous remark we

---

* Received April 10, 1961.

know that if we set $i = \operatorname{rank}(M)$, then $\operatorname{Ext}^i(M, R) \neq 0$. But it is well known that the hd M is the largest integer k such that $\operatorname{Ext}^k(M, R) \neq 0$ (see [2; VI, Exercise 9]). Therefore setting $h = k - i \geqq t$ we have the desired integers i and h.

On the other hand suppose that $\operatorname{Ext}^i{}_R(M, R) \neq 0$ and $\operatorname{Ext}^{i+h}{}_R(M, R) \neq 0$ with $h \geqq t$. Then it follows from what has been observed before that the rank $(M) \leqq i$ and that the hd $M \geqq i + h$. Therefore

$$d(M) = \operatorname{hd} M - \operatorname{rank} M \geqq h \geqq t.$$

We say that a noetherian ring R is regular if $R_\mathfrak{p}$ is a regular local ring for each prime ideal $\mathfrak{p}$ in R and $\dim R$ is finite.

THEOREM 2. *Let R be a regular ring, M a finitely generated, non-zero R-module. For each integer $t \geqq 0$, there exists an ideal $\mathfrak{c}_t \supset \mathfrak{a}(M)$ such that a prime ideal $\mathfrak{p} \supset \mathfrak{c}_t$ if and only if $d(M_\mathfrak{p}) \geqq t$.*

Proof. Let $\mathfrak{a}_i = \mathfrak{a}(\operatorname{Ext}^i{}_R(M, R))$ for all $i \geqq 0$. It is clear that each $\mathfrak{a}_i \supset \mathfrak{a}(M)$ and only a finite number of $\mathfrak{a}_i \neq R$ since only a finite number of $\operatorname{Ext}^i{}_R(M, R) \neq 0$. For each pair of integers $i \geqq 0$ and $h \geqq 0$ define $\mathfrak{b}(i, h) = (\mathfrak{a}_i, \mathfrak{a}_{i+h})$. It should be observed that only a finite number of the $\mathfrak{b}(i, h) \neq R$ and that each $\mathfrak{b}(i, h) \supset \mathfrak{a}(M)$.

For each integer $t \geqq 0$ define $\mathfrak{c}_t$ to be $\bigcap_{h \geqq t} \mathfrak{b}(i, h)$. The fact that $\mathfrak{c}_t \supset \mathfrak{a}(M)$ is trivial. Since $\bigcap_{h \geqq t} \mathfrak{b}(i, h)$ is really a finite intersection, we know that a prime ideal $\mathfrak{p}$ in R contains $\mathfrak{c}_t$ if and only if $\mathfrak{p} \supset \mathfrak{b}(i, b)$ for some i and some $h \geqq t$. But $\mathfrak{p} \supset \mathfrak{b}(i, h)$ if and only if $R_\mathfrak{p} \otimes \operatorname{Est}^i{}_R(M, R) \neq 0$ and $R_\mathfrak{p} \otimes \operatorname{Ext}^{i+h}{}_R(M, R) \neq 0$. Since $R_\mathfrak{p}$ is R-flat, it follows from [2; VI, Exercise 11] that $R_\mathfrak{p} \otimes \operatorname{Ext}^j{}_R(M, R) = \operatorname{Ext}^j{}_{R_\mathfrak{p}}(M_\mathfrak{p}, R_\mathfrak{p})$ for all j. Therefore applying Proposition 1, we have that $\mathfrak{p} \supset \mathfrak{c}_t$ if and only if $d(M_\mathfrak{p}) \geqq t$.

The following module theoretic version of Theorem 0 of [4] is an easy consequence of Theorem 2.

COROLLARY 3. *Let S be a noetherian ring which is a factor ring of a regular ring and let M be a finitely generated non-zero S-module. Then for each integer $t \geqq 0$, the set of all prime ideals $\mathfrak{p} \in \operatorname{Spec}(S)$ such that $d(M_\mathfrak{p}) \geqq t$ forms a closed set.*

Proof. Let $f \colon R \rightarrow S$ be the canonical ring epimorphism. Let $\mathfrak{p}$ be a prime ideal in S and $\mathfrak{P}$ the prime ideal $f^{-1}(\mathfrak{p})$ in R. Then it is well known that $\dim(M_\mathfrak{P}) = \dim(M_\mathfrak{p})$ and $\operatorname{codim}(M_\mathfrak{P}) = \operatorname{codim}(M_\mathfrak{p})$. Therefore we have that $d(M_\mathfrak{P}) = d(M_\mathfrak{p})$.

Now given an integer $t \geqq 0$ we have by Theorem 2 that there exists an ideal $\mathfrak{c}_t$ in R which contains $\mathfrak{a}(M)$, and therefore the $\operatorname{Ker} f$, such that a prime ideal $\mathfrak{P}$ in R contains $\mathfrak{c}_t$ if and only if $d(M_{\mathfrak{P}}) \geqq t$. It therefore follows by standard arguments that a prime ideal $\mathfrak{p}$ in S contains $f(\mathfrak{c}_t)$ if and only if $d(M_{\mathfrak{p}}) \geqq t$.

BRANDEIS UNIVERSITY.

REFERENCES.

[1] M. Auslander and D. Buchsbaum, "Homological dimension in local rings," *Transactions of the American Mathematical Society*, vol. 85 (1957), pp. 390-405.
[2] H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton, 1956.
[3] D. Rees, "The grade of an ideal or module," *Proceedings of the Cambridge Philosophical Society*, vol. 52 (1957), pp. 28-42.
[4] H. Hironaka, "A generalized theorem of Krull Seidenberg on parameterized algebras of finite type," *American Journal of Mathematics*, vol. 82 (1960), pp. 831-850.

ON THE UNION OF TWO SOLID TORI.*

By C. H. EDWARDS, JR.[1]

The Poincare conjecture asserts that, if M is a simply connected closed 3-manifold, then M is topologically the 3-sphere S^3. Now every simply connected triangulated closed 3-manifold is orientable (because the fundamental group of a non-orientable triangulated closed 3-manifold is infinite), and every orientable triangulated closed 3-manifold is the union of two (orientable) solid tori of the same positive genus, intersecting in their common boundary. But every 3-manifold can be triangulated ([2] and [4]). In order to prove the Poincare conjecture, it would therefore suffice to show that, for every positive integer n, every simply connected closed 3-manifold which can be expressed as the union of two genus n solid tori intersecting in their common boundary is a 3-sphere. The object here is to consider a special case of this situation. Evereything will be considered from the semilinear point of view.

A collection Σ of mutually disjoint simple closed curves on a 2-manifold T will be called *independent on* T if and only if T does not contain a punctured 2-sphere (a 2-sphere minus the sum of the interiors of a finite number of disjoint closed 2-cells), each of whose boundary components is a curve in Σ. The closed 3-manifold M will be called *n-proper* if and only if M can be expressed as the union of two solid tori B_1 and B_2 of genus $n > 0$, intersecting in their common boundary T, with T containing an independent collection of $n-1$ mutually disjoint simple closed curves, each of which is nullhomotopic in both B_1 and B_2. Notice that the union of two genus 1 solid tori intersecting in their common boundary is automatically 1-proper. A closed 3-manifold will be called *proper* if and only if it is n-proper for some positive integer n.

THEOREM 1. *If M is a simply connected proper closed 3-manifold, then M is a 3-sphere.*

For the purpose of an inductive proof, it will be convenient to consider Theorem 1 in the form of the following sequence $\{A_n\}_{n=1}^{\infty}$ of Lemmas:

LEMMA A_n. *If M is a simply connected n-proper closed 3-manifold, then M is a 3-sphere.*

---

* Receive July 17, 1961.

[1] This research was supported in part by the National Science Foundation, Grant 8239.

Since Lemma A_1 is known to be true (see [3, p. 31] for instance), the first step of the inductive proof is available.

By a *surface of type* (p, q) is meant a compact orientable 2-manifold of genus p with q boundary components.

LEMMA B_n. *Suppose that the simply connected bounded 3-manifold C is the union of two genus n solid tori B_1 and B_2, with $A = B_1 \cap B_2$ being a surface of type $(n, 1)$ common to the boundaries of B_1 and B_2. Suppose further that $\mathrm{Bd}\, B_2$ contains an independent collection of $n-1$ mutually disjoint simple closed curves, each of which is nullhomotopic in both B_1 and B_2. If Lemma A_n is true, then it follows that C is a closed 3-cell.*

Proof. If $D_i = \mathrm{Cl}(\mathrm{Bd}\, B_i - A)$ for $i = 1, 2$, then D_1 and D_2 are disks, and the boundary of C is the 2-sphere $K = D_1 \cup D_2$. Let Q be a 3-cell, and identify its boundary with K. Assuming a given triangulation of C, triangulate Q by projection of the given triangulation of K to an interior point of Q. Since $C \cap Q = K$ is simply connected, it follows that $\pi_1(C \cup Q)$ is the free product of the trivial groups $\pi_1(C)$ and $\pi_1(Q)$, so that $C \cup Q$ is simply connected. Since the 3-cell Q intersects the genus n solid torus B_1 in the subdisk D_1 of its boundary, $B_1 \cup Q$ is a solid torus of genus n. Hence $C \cup Q = (B_1 \cup Q) \cup B_2$, with $B_1 \cup Q$ and B_2 being genus n solid tori intersecting in their common boundary $A \cup D_2$, and it follows from the hypotheses that the simply connected closed 3-manifold $C \cup Q$ is n-proper. Hence Lemma A_n implies that $C \cup Q$ is a 3-sphere. The polyhedral 2-sphere K therefore separates $C \cup Q$ into two components whose closures C and Q are closed 3-cells [1], [4].

LEMMA 1. *Suppose that B is a bounded 3-manifold which can be imbedded semilinearly in S^3, and that the boundary T of B is a closed orientable surface of genus h. If $\pi_1(B, T) = 1$ then B is a solid torus of genus h.*

Proof. In the proof of Theorem 19.1 of [5] Papakyriakopoulos uses the Dehn lemma and the loop theorem to show that B can be constructed by adding h solid handles to a simply connected 3-manifold Q bounded by a 2-sphere. Since it may be assumed that Q is a polyhedral subset of S^3, Q is a 3-cell [1], so that B is a solid torus of genus h.

LEMMA 2. *Suppose that B is a solid torus of genus m and that D is a polyhedral disk with $\mathrm{Int}\, D \subset \mathrm{Int}\, B$, such that $\mathrm{Bd}\, D$ is a simple closed polygon on $T = \mathrm{Bd}\, B$ separating T into two components of genra h and k*

respectively, where $h+k=m$. *Then the closures of the two components of* $B-D$ *are solid tori of genera* h *and* k.

Proof. Let A_1 and A_2 be the components of $T-\mathrm{Bd}\,D$ of genera h and k respectively. Let $T_i=A_i\cup D$ and denote by B_i the closure of the component of $B-D$ bounded by T_i, $i=1,2$.

Since B is a solid torus, $\pi_1(B,T)=1$, so that the injection homomorphism $i^*:\pi_1(T)\to\pi_1(B)$ is onto, because of the exactness of the homotopy sequence of (B,T). It follows that the injection $j^*:\pi_1(T_1\cup T_2)\to\pi_1(B_1\cup B_2)$ is also onto. Since $T_1\cap T_2=B_1\cap B_2=D$ is simply connected $\pi_1(T_1\cup T_2)$ is the free product of $\pi_1(T_1)$ and $\pi_1(T_2)$, while $\pi_1(B_1\cup B_2)$ is the free product of $\pi_1(B_1)$ and $\pi_1(B_2)$. But clearly j^* restricted to $\pi_1(T_i)$ is the natural injection $j_i^*:\pi_1(T_i)\to\pi_1(B_i)$, $i=1,2$. Since j^* is onto, j_1^* and j_2^* are onto, so that $\pi_1(B_1,T_1)=\pi_1(B_2,T_2)=1$. Because the solid torus B can be imbedded in S^3, it follows from Lemma 1 that B_1 and B_2 are solid tori of genera h and k respectively.

LEMMA 3. *Suppose that the simply connected closed 3-manifold* M *is the union of two solid tori* B_1 *and* B_2 *intersecting in their common boundary* T. *If* e *is a simple closed curve on* T *which is nullhomotopic in both* B_1 *and* B_2, *then* e *is nullhomologous on* T.

Proof. The Dehn lemma [6] implies that e bounds disks D_1 and D_2 with $\mathrm{Int}\,D_i\subset\mathrm{Int}\,B_i$, $i=1,2$. Now suppose that e is not nullhomologous on T, i. e. that e does not separate T. If x and y are any two points of $M-(D_1\cup D_2)$, then x and y can clearly be joined to points p and q respectively in T by arcs missing $D_1\cup D_2$, while p and q can be joined by an arc on T missing e. It follows that the 2-sphere $D_1\cup D_2$ does not separate M. But this contradicts the fact that M is simply connected (since by Poincare duality it follows that the 2-dimensional homology group of M is trivial). Consequently e is nullhomologous on T.

LEMMA 4. *Suppose that* Σ *is an independent collection of mutually disjoint nullhomologous simple closed curves on the closed 2-manifold* T. *Let* e *be a curve in* Σ, *with* D *a disk bounded by* e *whose interior does not meet* T. *If* A *is one component of* $T-e$, *and* Σ^* *is the collection of those curves of* Σ *which are contained in* A, *then* Σ^* *is an independent collection of mutually disjoint nullhomologous simple closed curves on the closed 2-manifold* $T^*=A\cup D$.

Proof. If r is in Σ^* and if T^*-r were connected, then $T-r$ would be connected. Hence each curve in Σ^* is nullhomologous on T^*.

Now suppose that the collection Σ^* is not independent on T^*, so that T^* contains a punctured 2-sphere S whose boundary components $r_1, \cdots, r_k$ are simple closed curves in Σ^*. If S and D are disjoint, then S is contained by T, which contradicts the fact that the collection Σ is independent on T. Otherwise, if S contains D, then $S - \operatorname{Int} D$ is a punctured 2-sphere in T whose boundary components are the curves $r_1, \cdots, r_k$, e of Σ, which is again a contradiction. Therefore Σ^* is an independent collection of mutually disjoint nullhomologous simple closed curves on T^*.

LEMMA 5. *If Σ is an independent collection of $n-1$ mutually disjoint nullhomologous simple closed curves on a genus n closed orientable 2-manifold T, then Σ contains a curve e which separates T into two components A_1 and A_2 whose closures are surfaces of types $(1,1)$ and $(n-1,1)$ respectively.*

Proof. First it will be shown by induction that an independent collection of nullhomologous mutually disjoint simple closed curves on a genus n closed orientable 2-manifold can contain no more than $n-1$ elements. Certainly this is true in the case $n=1$, since each nullhomologous simple closed curve on a torus bounds a disk. Supposing it to be true for all genera less than n, let Σ' be an independent collection of nullhomologous simple closed curves on a genus n closed orientable 2-manifold T'. Let e be an element of Σ' separating T' into two components A_1' and A_2' of genera k and $n-k$ respectively, $1 \leqq k \leqq n-1$. Let D be a disk bounded by e not meeting T' in its interior, and denote by Σ_1' and Σ_2' the collections consisting of those curves of Σ contained in A_1' and A_2' respectively. By Lemma 4, Σ_1' and Σ_2' are independent collections of nullhomologous mutually disjoint simple closed curves on $A_1' \cup D$ and $A_2' \cup D$ respectively. Since $A_1' \cup D$ and $A_2' \cup D$ are closed orientable 2-manifolds of genera $k < n$ and $n-k < n$ respectively, Σ_1' and Σ_2' contain at most $k-1$ and $n-k-1$ elements respectively. But $\Sigma' = \Sigma_1' \cup \Sigma_2' \cup \{e\}$, so that Σ' contains at most $(k+1)+(n-k-1)+1 = n-1$ elements.

Now consider the independent collection Σ of $n-1$ mutually disjoint nullhomologous simple closed curves on the genus n closed orientable 2-manifold T. Let h be the least positive integer such that Σ contains a curve e separating T into two components A_1 and A_2 whose closures are surfaces of types $(h,1)$ and $(n-h,1)$ respectively, and let D be a disk bounded by e whose interior does not meet T. Because of the minimality of h, the genus $n-h$ closed orientable 2-manifold $D \cup A_2$ contains each of the $n-2$ elements of $\Sigma - \{e\}$, while by Lemma 4, $\Sigma - \{e\}$ is an independent collection of mutually disjoint nullhomologous simple closed curves on $D \cup A_2$. If $h > 1$, then $n-2 \geqq n-h$, which contradicts the result of the previous paragraph. Consequently $h=1$, so that the lemma is established.

Proof of Theorem 1. Since Lemma A_1 is known to be true, it suffices to show that Lemma A_n follows from Lemma A_{n-1}. Hence suppose that Lemma A_{n-1} is true and let M be a simply connected n-proper closed 3-manifold.

Then M is the union of two genus n solid tori B_1 and B_2, intersecting in their common boundary T, with T containing an independent collection Σ of $n-1$ mutually disjoint simple closed curves, each of which is nullhomotopic in both B_1 and B_2. Since M is simply connected, Lemma 3 implies that each curve in Σ is nullhomologous on T. By Lemma 5 there is therefore a curve e in Σ which separates T into two components A^1 and A^{n-1} whose closures are surfaces of types $(1,1)$ and $(n-1,1)$ respectively. Since e is nullhomotopic in both B_1 and B_2, the Dehn lemma [6] can be used to find disks D_1 and D_2 bounded by e with $\operatorname{Int} D_i \subset \operatorname{Int} B_i$, $i=1,2$. Denoting by $B_i{}^1$ and $B_i{}^{n-1}$ the closures of the components of $B_i - D_i$ bounded by $A^1 \cup D_i$ and $A^{n-1} \cup D_i$ respectively, Lemma 2 implies that $B_i{}^1$ and $B_i{}^{n-1}$ are solid tori of genera 1 and $n-1$ respectively, $i=1,2$.

Let $C^1 = B_1{}^1 \cup B_2{}^1$ and $C^{n-1} = B_1{}^{n-1} \cup B_2{}^{n-1}$. Then $M = C^1 \cup C^{n-1}$ with C^1 and C^{n-1} intersecting in the 2-sphere $D^1 \cup D_2$, so that the trivial group $\pi_1(M)$ is the free product of $\pi_1(C^1)$ and $\pi_1(C^{n-1})$. Therefore C^1 and C^{n-1} are simply connected. It is now easily verified, using Lemma 4, that C^1 and C^{n-1} satisfy the hypothesis of Lemmas B_1 and B_{n-1} respectively, and are therefore 3-cells. Consequently M is the union of two 3-cells intersecting in their common boundary and is therefore a 3-sphere. This completes the proof of Theorem 1.

UNIVERSITY OF TENNESSEE.

REFERENCES.

[1] J. W. Alexander, "On the subdivision of 3-space by a polyhedron," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 10 (1924), pp. 6-8.

[2] R. H. Bing, "Locally tame sets are tame," *Annals of Mathematics*, vol. 59 (1954), pp. 145-158.

[3] ———, "Necessary and sufficient conditions that a 3-manifold be S^3," *Annals of Mathematics*, vol. 68 (1958), pp. 17-37.

[4] E. E. Moise, "Affine structures in 3-manifolds V," *Annals of Mathematics*, vol. 56 (1952), pp. 96-114.

[5] C. D. Papakyriakopoulos, "On solid tori," *Proceedings of the London Mathematical Society* (3), vol. 7 (1957), pp. 281-299.

[6] ———, "On Dehn's lemma and the asphericity of knot," *Annals of Mathematics*, vol. 66 (1957), pp. 1-26.

PARTITIONS OF LARGE MULTIPARTITES.*

By M. M. Robertson.

1. Introduction. Within the past few years, several authors have obtained asymptotic formulae for the number of different partitions of a multipartite number. A multipartite number of order j is a j dimensional vector, the components of which are non-negative integers, and a partition of $(n_1, n_2, \cdots, n_j)$ is a solution of the vector equation

$$(1.1) \qquad \sum_k (n_{1k}, n_{2k}, \cdots, n_{kj}) = (n_1, n_2, \cdots, n_j)$$

in multipartites other than $(0, 0, \cdots, 0)$. Two partitions, which differ only in the order of the multipartite numbers on the left-hand side of (1.1), are regarded as identical. We denote by $p_1(n_1, \cdots, n_j)$ the number of different partitions of $(n_1, \cdots, n_j)$ and by $p_2(n_1, \cdots, n_j)$ the number of those partitions in which no part has a zero component. Also, we write $p_3(n_1, \cdots, n_j)$ for the number of partitions of $(n_1, \cdots, n_j)$ into different parts and $p_4(n_1, \cdots, n_j)$ for the number of partitions into different parts none of which has a zero component.

In 1918, Hardy and Ramanujan [1] gave a very full discussion of the case $j = 1$, but it is only within the last decade that much attention has been paid to the asymptotic evaluation of the $p_r(n_1, \cdots, n_j)$ for $j > 1$. Auluck [2] obtained a formula for $p_1(n_1, n_2)$ when n_2 is fixed and n_1 is large and Nanda [4] showed that this formula remains valid when n_2 is large, provided that $n_2 = o(n_1^{\frac{1}{4}})$. In a recent article [10], I evaluated $p_r(n_1, \cdots, n_j)$ for $r = 1$ and 3 when one particular n_l tends to infinity more rapidly than the fourth power of every other n_l by means of an extension of Nanda's method and I also obtained an asymptotic formula for $p_r(n_1, \cdots, n_j)$ for $r = 2$ and 4 when one particular n_l tends to infinity more slowly than the cube root of every other n_l.

In his article, Auluck also obtained a formula for $p_1(n_1, n_2)$, where n_1 and n_2 are large but of the same order of magnitude, i.e. the ratio n_1/n_2 is bounded above and below, and, under the same conditions, Wright [6] found asymptotic expressions for $p_r(n_1, n_2)$, where $r = 1$, 2, 3 and 4. In this article, Wright also gave without proof the first terms of an asymptotic formula for

---

* Received September 7, 1961.

$\log p_2(n_1, \cdots, n_j)$, where every n_l is of the same order of magnitude. Meinardus [3] had just previously published a paper in which he had found the first term of this formula for multipartites. Later, Wright [7] obtained asymptotic expressions for $p_r(n_1, n_2)$ which hold for $n_1^{\frac{1}{2}+\epsilon_1} < n_2 < n_1^{2-\epsilon_2}$, where $r = 1, 2, 3$ and 4 and ϵ_1 and ϵ_2 are any fixed positive numbers. This is a substantial relaxation of the restrictions imposed upon n_1 and n_2 in both [2] and [6]. In his article, I extend Wright's method to derive formulae for $p_r(n_1, \cdots, n_j)$ for $r = 1, 2, 3$ and 4 and $n_1 \cdots n_j < \bar{n}^{j+1-\epsilon_3}$, where $\bar{n} = \min n_l$ and ϵ_3 is any fixed positive number. These results, however, are not valid when $j = 1$.

2. Notation and definitions. Throughout this article, k, l, m, n, q, r, ν, K and N represent non-negative integers and j is used for an integer greater than unity. C is a positive number, not necessarily the same at each occurrence, which may depend upon any r, ϵ_l but not upon any n_l, x_l, y_l, θ_l, ξ_l, z_l. The numbers ϵ_l are positive numbers, to be thought of as small. Except when there is a statement to the contrary, the symbols $\sim$, $o(\ )$ and $O(\ )$ refer to the passage of the n_l to infinity.

Before we can state the theorem which embodies the results of this paper, we require several definitions. We put

$$\lambda_1(\nu_1, \cdots, \nu_j) = \lambda_2(\nu_1, \cdots, \nu_j) = (1 - e^{-\nu_1 x_1 - \cdots - \nu_j x_j})^{-1}$$

and

$$\lambda_3(\nu_1, \cdots, \nu_j) = \lambda_4(\nu_1, \cdots, \nu_j) = 1 + e^{-\nu_1 x_1 - \cdots - \nu_j x_j},$$

where $R(x_l) > 0$ for $1 \leqq l \leqq j$. Then we write

$$f_r(x_1, \cdots, x_j) = \prod_{\nu_1, \cdots, \nu_j} \lambda_r(\nu_1, \cdots, \nu_j), \tag{2.1}$$

where, for $r = 2$ and 4, $\nu_1, \cdots, \nu_j$ each take all positive integral values, while, for $r = 1$ and 3, $\nu_1, \cdots, \nu_j$ each take all non-negative integral values except $\nu_1 = \cdots = \nu_j = 0$. If we put $p_r(0, 0, \cdots, 0) = 1$, we can easily verify that

$$f_r(x_1, \cdots, x_j) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_j=0}^{\infty} p_r(n_1, \cdots, n_j) e^{-n_1 x_1 - \cdots - n_j x_j}. \tag{2.2}$$

When $0 < |t| < 2\pi$,

$$\frac{t}{e^t - 1} = 1 - \tfrac{1}{2}t + \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \frac{B_\nu}{(2\nu)!} t^{2\nu},$$

where the B_ν are the Bernoulli numbers defined by

$$(2\pi)^{2\nu} B_\nu = 2(2\nu)!\, \zeta(2\nu).$$

We write x^* for the x_l with the greatest modulus, i. e. $|x^*| = \max |x_l|$ where $1 \leqq l \leqq j$, and we put $L = x_1 \cdots x_j$. Therefore, when $0 < t < 2\pi/|x^*|$,

$$t^{-1} \prod_{l=1}^{j} (e^{x_l t} - 1)^{-1} = L^{-1} t^{-j-1} \prod_{l=1}^{j} \{1 - \tfrac{1}{2} x_l t + \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \frac{B_\nu}{(2\nu)!} (x_l t)^{2\nu}\}.$$

Since every infinite series within the latter product is absolutely convergent, the product can be arranged as an infinite series in ascending powers of t. Therefore,

$$(2.3) \qquad t^{-1} \prod_{l=1}^{j} (e^{x_l t} - 1)^{-1} = L^{-1} t^{-j-1} \sum_{m=0}^{\infty} R_m t^m,$$

where $R_m = R_m(x_1, \cdots, x_j)$ is a symmetric homogeneous polynomial of degree m in the x_l. It is easily seen that, in particular, $R_0 = 1$ and $R_1 = -\frac{1}{2} \sum_{l=1}^{j} x_l$.

In [8], Wright showed that, when $|\arg x_l| < \frac{1}{2}\pi - \epsilon_4$ for any fixed positive number ϵ_4,

$$(2.4) \qquad L \log f_2(x_1, \cdots, x_j) = H + \sum_{\substack{m=0 \\ m \neq j}}^{2K+j-1} \zeta(j + 1 - m) R_m + O(x^{*2K+j})$$

as $x^* \to 0$ for any $K > 0$, where

$$(2.5) \qquad \begin{aligned} H &= H(x_1, \cdots, x_j) \\ &= \gamma R_j + \int_0^\infty \{L t^{-1} \prod_{l=1}^{j} (e^{x_l t} - 1)^{-1} - \sum_{m=0}^{j-1} R_m t^{m-j-1} - R_j (e^t - 1)^{-1}\} dt \end{aligned}$$

and

$$\gamma = \lim_{n \to \infty} (\sum_{q=1}^{n} q^{-1} - \log n).$$

We write $\bar{L} = z_1 \cdots z_{j-1}$ and $\bar{R}_m = R_m(z_0, z_1, \cdots, z_{j-1})$, where $z_0 = 1$, and we transform the integral on the right-hand side of (2. 5) by putting $u = x_k t$ and $z_{kl} = x_k^{-1} x_l$ for $1 \leqq k \leqq j$ and $1 \leqq l \leqq j$. Then, since

$$(2.6) \qquad \begin{aligned} &\int_0^\infty \{(e^t - 1)^{-1} - x_k (e^{x_k t} - 1)^{-1}\} dt \\ &\qquad = [\log\{(1 - e^{-t})/(1 - e^{-x_k t})\}]_0^\infty = \log x_k, \end{aligned}$$

it follows that

$$(2.7) \qquad H = \gamma R_j + j^{-1} \{L \sum_{k=1}^{j} I(z_{k1}, \cdots, z_{k,k-1}, z_{k,k+1}, \cdots, z_{kj}) - R_j \log L\},$$

where

$$(2.8) \qquad I = I(z_1, \cdots, z_{j-1}) = \int_0^\infty h(u : z_1, \cdots, z_{j-1}) du$$

and

$$(2.9)\quad h(u:z_1,\cdots,z_{j-1})=u^{-1}\prod_{l=0}^{j-1}(e^{z_l u}-1)^{-1}-L^{-1}\{\sum_{m=0}^{j-1}R_m u^{m-j-1}+R_j(e^u-1)^{-1}\}.$$

Since $R(x_l)>0$ for $1\leqq l\leqq j$, the integral on the right-hand side of (2.5) converges. Also, from (2.8) and (2.9), every

$$I(z_{k1},\cdots,z_{k,k-1},z_{k,k+1},\cdots,z_{kj})$$

converges for $1\leqq k\leqq j$, provided that $R(z_{kl})\neq 0$ for $1\leqq k\leqq j$ and $1\leqq l\leqq j$. To prove the theorem of the next section, we require the values of the different $I(z_{k1},\cdots,z_{k,k-1},z_{k,k+1},\cdots,z_{kj})$ only when $\arg x_l=o(1)$ for $1\leqq l\leqq j$ and so, as $R(z_{kl})>0$ for all k and l, every $I(z_{k1},\cdots,z_{k,k-1},z_{k,k+1},\cdots,z_{kj})$ converges in this case.

From the definitions of the λ_r it is easily seen that

$$\lambda_3(\nu_1,\cdots,\nu_j)=\lambda_1(\nu_1,\cdots,\nu_j)/\lambda_1(2\nu_1,\cdots,2\nu_j),$$

$$\lambda_4(\nu_1,\cdots,\nu_j)=\lambda_2(\nu_1,\cdots,\nu_j)/\lambda_2(2\nu_1,\cdots,2\nu_j)$$

and hence, from (2.1),

$$f_3(x_1,\cdots,x_j)=f_1(x_1,\cdots,x_j)/f_1(2x_1,\cdots,2x_j),$$

$$f_4(x_1,\cdots,x_j)=f_2(x_1,\cdots,x_j)/f_2(2x_1,\cdots,2x_j).$$

Therefore,

$$(2.10)\quad \log f_3(x_1,\cdots,x_j)=\log f_1(x_1,\cdots,x_j)-\log f_1(2x_1,\cdots,2x_j)$$

and

$$(2.11)\quad \log f_4(x_1,\cdots,x_j)=\log f_2(x_1,\cdots,x_j)-\log f_2(2x_1,\cdots,2x_j).$$

It can also be seen from (2.1) that

$$\begin{aligned}&\log f_1(x_1,\cdots,x_j)\\ (2.12)\quad &=\log f_2(x_1,\cdots,x_j)+\sum_{l=1}^{j}\log f_2(x_1,\cdots,x_{l-1},x_{l+1},\cdots,x_j)\\ &\quad+\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}\log f_2(x_1,\cdots,x_{l_1-1},x_{l_1+1},\cdots,x_{l_2-1},x_{l_2+1},\cdots,x_j)\\ &\quad+\cdots+\sum_{l=1}^{j}\log f_2(x_l)\end{aligned}$$

and

$$\log f_3(x_1,\cdots,x_j)$$

$$
\begin{aligned}
(2.13)\qquad &= \log f_4(x_1,\cdots,x_j) + \sum_{l=1}^{j} \log f_4(x_1,\cdots,x_{l-1},x_{l+1},\cdots,x_j) \\
&\quad + \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} \log f_2(x_1,\cdots,x_{l_1-1},x_{l_1+1},\cdots,x_{l_2-1},x_{l_2+1},\cdots,x_j) \\
&\quad + \cdots + \sum_{l=1}^{j} \log f_2(x_l).
\end{aligned}
$$

We write

$$T_{2m} = T_{2m}(x_1,\cdots,x_j) = \zeta(j+1-m) R_m(x_1,\cdots,x_j)$$

for all $m \neq j$,

$$T_{2j} = T_{2j}(x_1,\cdots,x_j) = \gamma R_j(x_1,\cdots,x_j)$$

and

$$\tau_{2j\nu} = \tau_{2j\nu}(x_1,\cdots,x_j) = -j^{-1} R_j(x_1,\cdots,x_j)$$

for $1 \leqq \nu \leqq j$. Next, we write

$$
\begin{aligned}
T_{1m} &= T_{1m}(x_1,\cdots,x_j) \\
&= T_{2m}(x_1,\cdots,x_j) + \sum_{l=1}^{j} x_l T_{2,m-1}(x_1,\cdots,x_{l-1},x_{l+1},\cdots,x_j) \\
&\quad + \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} x_{l_1} x_{l_2} T_{2,m-2}(x_1,\cdots,x_{l_1-1},x_{l_1+1},\cdots,x_{l_2-1},x_{l_2+1},\cdots,x_j) \\
&\quad + \cdots + \sum_{l=1}^{j} x_1 \cdots x_{l-1} x_{l+1} \cdots x_j T_{2,m-j+1}(x_l)
\end{aligned}
$$

for all m, where $T_{2,-k} = 0$ for $k > 0$, and

$$
\begin{aligned}
\tau_{1j\nu} &= \tau_{1j\nu}(x_1,\cdots,x_j) \\
&= \tau_{2j\nu}(x_1,\cdots,x_j) + \sum_{\substack{l=1 \\ l\neq\nu}}^{j} x_l \tau_{2,j-1,\nu}(x_1,\cdots,x_{l-1},x_{l+1},\cdots,x_j) \\
&\quad + \sum_{\substack{l_1=1 \\ l_1\neq\nu}}^{j-1} \sum_{\substack{l_2=l_1+1 \\ l_2\neq\nu}}^{j} x_{l_1} x_{l_2} \tau_{2,j-2,\nu}(x_1,\cdots,x_{l_1-1},x_{l_1+1},\cdots,x_{l_2-1},x_{l_2+1},\cdots,x_j) \\
&\quad + \cdots + x_1 \cdots x_{\nu-1} x_{\nu+1} \cdots x_j \tau_{21\nu}(x_\nu).
\end{aligned}
$$

Then we put

$$T_{3m} = T_{3m}(x_1,\cdots,x_j) = (1-2^{m-j}) T_{1m}(x_1,\cdots,x_j)$$

and

$$T_{4m} = T_{4m}(x_1,\cdots,x_j) = (1-2^{m-j}) T_{2m}(x_1,\cdots,x_j)$$

for all $m \neq j$ and

$$T_{3j} = T_{3j}(x_1, \cdots x_j)$$
$$= (\log 2)\{R_j(x_1, \cdots, x_j) + \sum_{l=1}^{j} x_l R_{j-1}(x_1, \cdots, x_{l-1}, x_{l+1}, \cdots, x_j)$$
$$+ \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} x_{l_1} x_{l_2} R_{j-2}(x_1, \cdots, x_{l_1-1}, x_{l_1+1}, \cdots, x_{l_2-1}, x_{l_2+1}, \cdots, x_j)$$
$$+ \cdots + \sum_{l=1}^{j} x_1, \cdots, x_{l-1}, x_{l+1}, \cdots, x_j R_1(x_l)\}$$

and

$$T_{4j} = T_{4j}(x_1, \cdots, x_j) = (\log 2) R_j(x_1, \cdots, x_j).$$

We write also

$$\tau_{3j\nu} = \tau_{4j\nu} = \bar{\tau}_{30} = \bar{\tau}_{40} = 0,$$

$$\bar{\tau}_{20} = \bar{\tau}_{20}(x_1, \cdots, x_j) = j^{-1} \sum_{k=1}^{j} I(z_{k1}, \cdots, z_{k,k-1}, z_{k,k+1}, \cdots, z_{kj})$$

and

$$\bar{\tau}_{10} = \bar{\tau}_{10}(x_1, \cdots, x_j)$$
$$= -\tfrac{1}{2} j(\log 2\pi - \gamma) + \sum_{\nu=2}^{j} \nu^{-1} \sum_{\nu} \sum_{k=1}^{\nu} I(z_{l_k l_1}, \cdots, z_{l_k l_{k-1}}, z_{l_k l_{k+1}}, \cdots, z_{l_k l_\nu}),$$

where the sum $\sum_{\nu}$ is taken over the $C_\nu{}^j$ different ways of choosing ν different integers $l_1, \cdots, l_\nu$ from $1, 2, \cdots, j$. It can easily be verified that

$$(2.14) \qquad T_{10} = T_{20} = \zeta(j+1), \qquad T_{30} = T_{40} = (1 - 2^{-j})\zeta(j+1)$$

and

$$T_{r,2m+j+1} = 0$$

for $m > 0$, since $\zeta(-2m) = 0$. We now deduce from (2.4), (2.7), (2.10), (2.11), (2.12) and (2.13) that

$$(2.15) \qquad L \log f_r(x, \cdots, x_j) = \sum_{m=0}^{2K+j-1} T_{rm} + \sum_{\nu=1}^{j} \tau_{rj\nu} \log x_\nu + L\bar{\tau}_{r0} + O(x^{*2K+j})$$

as $x^* \to 0$.

For $1 \leqq l \leqq j$ we put

$$x_l = y_l + i\theta_l = y_l(1 + i\xi_l),$$

where $y_l > 0$ and $-\pi \leqq \theta_l \leqq \pi$, and for all r we let

$$y_l \sim \mu_{rl}, \tag{2.16}$$

where the μ_{rl} are defined by

$$\mu_{rl} = n_l^{-1}(T_{r0}n_1 \cdots n_j)^{1/(j+1)}. \tag{2.17}$$

We write $L^R = y_1 \cdots y_j$, $T_{rm}{}^R = T_{rm}(y_1, \cdots, y_j)$, $\tau_{rj\nu}{}^R = \tau_{rj\nu}(y_1, \cdots, y_j)$ and $\bar{\tau}_{r0}{}^R = \bar{\tau}_{r0}(y_1, \cdots, y_j)$. From (4.3), which we prove later, we see that a fixed integer K can be chosen such that K is the smallest integer for which $\mu_{rl}{}^{2K+j} = o(\mu_{r1} \cdots \mu_{rj})$ for $1 \leqq l \leqq j$ and we deduce from (2.15) that

$$\begin{aligned} \log f_r(y_1, \cdots, y_j) \\ = L^{R^{-1}}\Big(\sum_{m=0}^{2K+j-1} T_{rm}{}^R + \sum_{\nu=1}^{j} \tau_{rj\nu}{}^R \log y_\nu\Big) + \bar{\tau}_{r0}{}^R + o(1). \end{aligned} \tag{2.18}$$

Also, we define χ_{rl} for $1 \leqq l \leqq j$ by

$$\begin{aligned} \chi_{rl} = \chi_{rl}(y_1, \cdots, y_j) \\ = -L^R T_{r0}{}^{-1} y_l \frac{\partial}{\partial y_l}\{L^{R^{-1}}\Big(\sum_{m=0}^{2K+j-1} T_{rm}{}^R + \sum_{\nu=1}^{j} \tau_{rj\nu}{}^R \log y_\nu\Big) + \bar{\tau}_{r0}{}^R\}. \end{aligned} \tag{2.19}$$

When $r = 3$ or 4, we can see that, since $\tau_{rj\nu} = \bar{\tau}_r = 0$, the terms of χ_{rl} form a polynomial in the y_l with unity as the term of zero degree. We shall see later that, when $r = 1$ or 2, in addition to a similar polynomial in the y_l, the terms of χ_{rl} which are not $o\{\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_5}\}$ for some fixed positive number ϵ_5 may also include terms involving definite integrals of a type similar to, but of a more general nature than, the integral I defined by (2.8) and (2.9).

3. Statement and proof of the theorem.

Employing the definitions of the previous section, we now prove the

THEOREM. *If, for $1 \leqq l \leqq j$ $(j > 1)$, every n_l tends to infinity subject to the condition that*

$$n_1 \cdots n_j < n_l{}^{j+1-\epsilon_8} \tag{3.1}$$

for any fixed positive number ϵ_8, then

$$p_r(n_1, \cdots, n_j) \sim \exp \Xi_r,$$

where

$$\begin{aligned} \Xi_r = \sum_{l=1}^{j} n_l y_l + L^{R^{-1}}\Big(\sum_{m=0}^{2K+j-1} T_{rm}{}^R + \sum_{\nu=1}^{j} \tau_{rj\nu}{}^R \log y_\nu\Big) \\ + \bar{\tau}_{r0}{}^R + \tfrac{1}{2}(j+2)\log L^R - \tfrac{1}{2}j \log(2\pi T_{r0}) - \tfrac{1}{2}\log(j+1) \end{aligned} \tag{3.2}$$

and the y_l *are functions of* $\mu_{r1}, \cdots, \mu_{rj}$ *which satisfy* (2.15) *and the equations*

$$n_l y_l L^R = T_{r0} \chi_{rl} + o\{(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_5}\} \tag{3.3}$$

for $1 \leqq l \leqq j$ *and any fixed positive number* ϵ_5.

We first choose ϵ_5 to satisfy

$$0 < 2(j+1)\epsilon_5 < \epsilon_3. \tag{3.4}$$

Next, we define η and U by

$$\eta = (\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}-\epsilon_5}, \qquad U = \eta (T_{r0} L^{R-1})^{\frac{1}{2}} \sim C(n_1 \cdots n_j)^{\epsilon_5/(j+1)}.$$

From (2.2), we obtain

$$p_r(n_1, \cdots, n_j) = (2\pi)^{-j} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{F_r(x_1, \cdots, x_j)} d\theta_1 \cdots d\theta_j,$$

where

$$F_r(x_1, \cdots,) = \sum_{l=1}^{j} n_l x_l + \log f_r(x_1, \cdots, x_j), \tag{3.5}$$

and therefore,

$$p_r(n_1, \cdots, n_j) = I_1 + I_2, \tag{3.6}$$

where

$$I_1 = L^R (2\pi)^{-j} \int \cdots \int_{\sum_{l=1}^{j} |\xi_l| \geqq \eta} e^{F_r(x_1, \cdots, x_j)} d\xi_1 \cdots d\xi_j$$

and

$$I_2 = L^R (2\pi)^{-j} \int \cdots \int_{\sum_{l=1}^{j} |\xi_l| \leqq \eta} e^{F_r(x_1, \cdots, x_j)} d\theta_1 \cdots d\theta_j.$$

From the formula immediately preceding relation (7) of [10], we easily deduce that

$$\log f_r(x_1, \cdots, x_j) = \sum_{k=1}^{\infty} k^{-1} \prod_{l=1}^{j} (e^{kx_l} - 1)^{-1}$$

and, from this result and (2.10), (2.11), (2.12) and (2.13), the following lemma can be proved by an argument similar to that used to prove Lemma 1 of [6].

LEMMA 1. *If* $\sum_{l=1}^{j} |\xi_l| \geqq \eta$, *then*

$$|\log f_r(x_1, \cdots, x_j)| \leqq \log f_r(y_1, \cdots, y_j) - CU^2.$$

We shall prove later the following lemmas.

LEMMA 2. *If* $\sum_{l=1}^{j} |\xi_l| \leqq \eta$, *then*

$$F_r(x_1,\cdots,z_j)+F_r(y_1,\cdots,y_j)$$
$$=-T_{r0}L^{R-1}\Big(\sum_{l=1}^{j}\xi_l^2+\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}\xi_{l_1}\xi_{l_2}\Big)+o(1).$$

LEMMA 3. *It is always possible to choose* $y_1,\cdots,y_j$ *to satisfy* (2.15) *and* (3.3).

LEMMA 4.

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\Big(-\sum_{l=1}^{j}u_l^2-\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}u_{l_1}u_{l_2}\Big)du_1\cdots du_j=(j+1)^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}j}.$$

It follows from (3.5) and Lemma 1 that

$$|I_2|<e^{F_r(y_1,\cdots,y_j)-CU^2}. \tag{3.7}$$

Also, from Lemma 2,

$$I_1\sim L^R(2\pi)^{-j}e^{F_r(y_1,\cdots,y_j)}$$
$$\underset{\sum_{l=1}^{j}|\xi_l|\leqq\eta}{\int\cdots\int}\exp\Big\{-T_{r0}L^{R-1}\Big(\sum_{l=1}^{j}\xi_l^2+\sum_{l_1=1}^{l=1}\sum_{l_2=l_1+1}^{j}\xi_{l_1}\xi_{l_2}\Big)\Big\}d\xi_1\cdots d\xi_j.$$

If we transform the variables in the latter integral by putting

$$\xi_l=(L^RT_{r0}^{-1})^{\frac{1}{2}}u_l$$

for $1\leqq l\leqq j$, we obtain

$$I_1\sim L^{R\frac{1}{2}(j+2)}(4\pi^2T_{r0})^{-\frac{1}{2}j}e^{F_r(y_1,\cdots,y_j)}$$
$$\times\underset{\sum_{l=1}^{j}|u_l|\leqq U}{\int\cdots\int}\exp\Big(-\sum_{l=1}^{j}u_l^2-\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}u_{l_1}u_{l_2}\Big)du_1\cdots du_j.$$

Then, since the integrand is positive everywhere,

$$\int_{-j^{-1}U}^{j^{-1}U}\cdots\int_{-j^{-1}U}^{j^{-1}U}<\underset{\sum_{l=1}^{j}|u_l|\leqq U}{\int\cdots\int}$$
$$\int_{-U}^{U}\cdots\int_{-U}^{U}\exp\Big(-\sum_{l=1}^{j}u_l^2-\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}u_{l_1}u_{l_2}\Big)du_1\cdots du_j$$

and (3.2) follows from (2.18), (3.5), (3.6), (3.7) and Lemma 4.

In any practical application of the theorem, it is necessary to determine the y_l as functions of the μ_{rl} and hence the n_l for $1 \leqq l \leqq j$ from the equations (3.3) but, as this is also required for the proof of Lemma 3, we defer the problem to the following section. When the y_l have been evaluated, all the terms in (3.2) can be calculated with the exception of $\bar{\tau}_{10}{}^R$ and $\bar{\tau}_{20}{}^R$. In order to calculate these latter terms, we must evaluate $I(z_1, \cdots, z_{j-1})$ and this problem is discussed fully in [11]. We could add any terms to y_l which are $o\{\mu_{rl}(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_5}\}$ for $1 \leqq l \leqq j$ without affecting the truth of (3.3). However, since from (2.19), (3.2) and (3.3),

$$\frac{\partial \Xi_r}{\partial y_l} = o\{\mu_{rl}^{-1}(\mu_{r1} \cdots \mu_{rj})^{-\frac{1}{2}+\epsilon_5}\}$$

any change in y_l of order $o\{\mu_{rl}(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_5}\}$ will produce a change in Ξ_r of order $o\{\mu_{r1} \cdots \mu_{rj})^{2\epsilon_5}\} = o(1)$. The values of the y_l, therefore, calculated from (3.3) to within an order $o\{\mu_{rl}(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_5}\}$ for any positive number ϵ_5, determine completely the formula (3.2) of the theorem.

$$\log p_r(n_1, \cdots, n_j) \sim (j+1)(T_{r0} n_1 \cdots n_j)^{1/(j+1)}$$

is an immediate consequence and this shows that the result of Meinardus [3], which was given for every n_l of the same order of magnitude, holds under a considerable relaxation of the conditions imposed upon the n_l. Wright [7] stated the first few terms in the asymptotic expansion of $\log p_2(n_1, \cdots, n_j)$, when every n_l is of the same order of magnitude, and this also can be easily deduced from the theorem. Although, for $r = 1$ or 2, the asymptotic formulae for $p_r(n_1, \cdots, n_j)$ involve the evaluation of definite integrals, formulae for $\dfrac{p_r(kn_1, \cdots, kn_j)}{p_r(n_1, \cdots, n_j)}$, where k is any fixed positive integer, are free from these integrals and can be expressed in terms of elementary functions.

4. Proof of Lemmas 2 and 3. In both the proof of Lemma 2 and the proof of Lemma 3, we make use of properties of the definite integral I which are proved in [11]. We begin the proof of Lemma 2 by defining the linear operator Ω by

$$\Omega g(x_1, \cdots, x_j) \\ = g(x_1, \cdots, x_j) - g(y_1, \cdots, y_j) - \sum_{l=1}^{j} i\xi_l y_l \frac{\partial}{\partial y_l} g(y_1, \cdots, y_j),$$

where $g(x_1, \cdots, x_j)$ is any function of the variables $x_1, \cdots, x_j$ with continuous first order partial derivatives. First, we prove that

$$(4.1) \qquad \Omega F_r{}^{\ddagger}(x_1, \cdots, x_j) = -T_{r0} L^{R-1}\Big(\sum_{l=1}^{j} \xi_l{}^2 + \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} \xi_{l_1}\xi_{l_2}\Big) + o(1),$$

where

$$(4.2)\qquad F_r^*(x_1,\cdots,x_j)=\sum_{l=1}^{j}n_lx_l+L^{-1}\Big(\sum_{m=0}^{2K+j-1}T_{rm}+\sum_{\nu=1}^{j}\tau_{rj\nu}\log x_\nu\Big)+\bar{\tau}_{r0}.$$

Trivial calculations show that

$$\Omega\Big(\sum_{l=1}^{j}n_lx_l\Big)=0,$$

$$\Omega(L^{-1})=-L^{R-1}\Big(\sum_{l=1}^{j}\xi_l^2+\sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}\xi_{l_1}\xi_{l_2}\Big)+O(\eta^3L^{R-1}),$$

$$\Omega(L^{-1}\tau_{rj\nu}\log x_\nu)=O(\eta^2L^{R-1}\tau_{rj\nu}{}^R\log y_\nu)$$

and, for all $m>0$,

$$\Omega(L^{-1}T_{rm})=O(\eta^2L^{R-1}T_{rm}{}^R).$$

Now,

$$\eta^3L^{R-1}=O\{(\mu_{r1}\cdots\mu_{rj})^{\frac{1}{2}-3\epsilon_5}\}=o(1)$$

from (3.4). Putting $\mu_r^*=\max\mu_{rl}$, $\bar{\mu}_r=\min\mu_{rl}$, $n^*=\max n_l$ and $\bar{n}=\min n_l$, we see from (2.17) that

$$\mu_r^*=\bar{n}^{-1}(T_{r0}n_1\cdots n_j)^{1/(j+1)},\quad \bar{\mu}_r=n^{*-1}(T_{r0}n_1\cdots n_j)^{1/(j+1)}.$$

From (3.1),

$$\bar{n}^j\leqq n_1\cdots n_j<\bar{n}^{j+1-\epsilon_3},\qquad n^*<\bar{n}^{2-\epsilon_3},$$

and therefore,

$$C\bar{n}^{-1+j\epsilon_3/(j+1)}<\bar{\mu}_r\leqq\mu_r^*<C\bar{n}^{-\epsilon_3/(j+1)}.$$

Hence, we have

$$(4.3)\qquad \bar{\mu}_r\leqq\mu_r^*<\bar{\mu}_r{}^{\epsilon_3/(j+1)}$$

and

$$\log\bar{\mu}_r=O(\log\mu_r^*).$$

Also, both $\eta^2L^{R-1}|T_{rm}{}^R|$ for $m>0$ and $\eta^2L^{R-1}|\tau_{rj\nu}{}^R\log y_\nu|$ for $1\leqq\nu\leqq j$ are less than

$$C\mu_r^*(\mu_{r1}\cdots\mu_{rj})^{-2\epsilon_5}=C\bar{n}^{-1}(n_1\cdots n_j)^{(2\epsilon_5+1)/(j+1)}$$

$$<C\bar{n}^{2\epsilon_5-\epsilon_3(2\epsilon_5+1)/(j+1)}=o(1)$$

by (3.4). We have, therefore,

$$\Omega(L^{-1}) = -L^{R-1}\Big(\sum_{l=1}^{j}\xi_l^2 + \sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j}\xi_{l_1}\xi_{l_2}\Big) + o(1),$$

$$\Omega(L^{-1}\tau_{rj\nu}\log x_\nu) = o(1)$$

and

$$\Omega(L^{-1}T_{rm}) = o(1)$$

for all $m > 0$.

We put $y_m = \max y_l$. Then we deduce from Taylor's Theorem that

$$(e^{z_{ml}u} - 1)^{-1} = (e^{y_l u\{1+i(\xi_l-\xi_m)+O(\eta^2)\}/y_m} - 1)^{-1}$$

$$= (e^{y_l u/y_m} - 1)^{-1} + \{i(\xi_m - \xi_l) + O(\eta^2)\}(y_l/y_m)e^{y_l u/y_m}(e^{y_l u/y_m} - 1)$$
$$+ O\{\eta^2(e^{y_l u/2y_m} - 1)^{-1}\}$$

for all $u > 0$. It follows that

$$\Omega\{u^{-1}\prod_{l=1}^{j}(e^{z_{ml}u} - 1)^{-1}\} = O\{\eta^2 u^{-1}\prod_{l=1}^{j}(e^{y_l u/2y_m} - 1)^{-1}\}$$
$$= O\{\eta^2 y_m^{j}(y_1\cdots y_j)^{-1}u^{-j-1}\}$$

for all $u > 0$ and therefore,

$$\Omega\{\int_{\frac{1}{2}\pi}^{\infty} u^{-1}\prod_{l=1}^{j}(e^{z_{ml}u} - 1)^{-1}\} = O(\eta^2\mu_r^{*j}/\mu_{r1}\cdots\mu_{rj}).$$

We can easily show that

$$\Omega\{(z_{m1}\cdots z_{mj})^{-1}\int_{\frac{1}{2}\pi}^{\infty}\{\sum_{\nu=0}^{j-1}R_\nu(z_{m1},\cdots,z_{mj})e^{\nu-j-1}$$
$$+ R_j(z_{m1},\cdots,z_{mj})(e^{u} - 1)^{-1}\}du\} = O(\eta^2\mu_r^{*j}/\mu_{r1}\cdots\mu_{rj}).$$

Also, we have

$$\sum_{\nu=0}^{\infty} y_m^{-\nu}\,|\,R_\nu(y_1,\cdots,y_j)|\,\pi^\nu \leqq \prod_{l=1}^{j}\Big(1 + \tfrac{1}{2}\pi + \sum_{k=1}^{\infty}\frac{B_k\pi^{2k}}{(2k)!}\Big) = C$$

and so,

$$y_m^{-\nu}\,|\,R_\nu(y_1,\cdots,y_j)\,| < C\pi^{-\nu}.$$

Hence, since for $0 \leqq u \leqq \frac{1}{2}\pi$,

$$h(u : z_{m1},\cdots,z_{m,m-1},z_{m,m+1},\cdots,z_{mj})$$
$$= (z_{m1}\cdots z_{mj})^{-1}\{\sum_{\nu=j+1}^{\infty}R_\nu(z_{m1},\cdots,z_{mj})u^{\nu-j-1}$$
$$+ R_j(z_{m1},\cdots,z_{mj})(u^{-1} - (e^{u} - 1)^{-1})\},$$

we obtain

$$\Omega\{\int_0^{\frac{1}{2}\pi} h(u: z_{m1}, \cdots, z_{m,m-1}, z_{m,m+1}, \cdots, z_{mj})du\}$$

$$= O\{\eta^2 y_m{}^j (y_1 \cdots y_j)^{-1} \sum_{\nu=j}^{\infty} 2^{j-\nu}\}$$

$$= O(\eta^2 \mu_r{}^{*j}/\mu_{r1} \cdots \mu_{rj}).$$

Therefore,

$$\Omega I(z_{m1}, \cdots, z_{m,m-1}, z_{m,m+1}, \cdots, z_{mj}) = O(\eta^2 \mu_r{}^{*j}/\mu_{r1} \cdots \mu_{rj})$$

$$= O(\bar{n}^{2\epsilon_5 - \epsilon_3(2\epsilon_5 + j)/(j+1)}) = o(1).$$

Now, for $1 \leqq l \leqq j$, we have

$$I(z_{l1}, \cdots, z_{l,l-1}, z_{l,l+1}, \cdots, z_{lj})$$

$$= I(z_{m1}, \cdots, z_{m,m-1}, z_{m,m+1}, \cdots, z_{mj}) - L^{-1}R_j(x_1, \cdots, x_j) \log z_{lm}$$

from (3.4) of [11] and therefore, since a trival calculation shows that

$$\Omega(L^{-1}R_j \log z_{lm}) = o(1),$$

$$\Omega I(z_{l1}, \cdots, z_{l,l-1}, z_{l,l+1}, \cdots, z_{lj}) = o(1).$$

It follows that, for $r = 1$ and 2,

$$\Omega \bar{\tau}_{r0} = o(1).$$

This completes the proof of (4.1). From (4.2) and (2.19),

$$y_l \frac{\partial}{\partial y_l} F_r{}^*(y_1, \cdots, y_j) = n_l y_l - L^{R-1} T_{r0} \chi_{rl}$$

and hence, from (3.3), we have

$$i\xi_l y_l \frac{\partial}{\partial y_l} F_r{}^*(y_1, \cdots, y_j) = o(1)$$

for $1 \leqq l \leqq j$. Therefore,

$$F_r{}^*(x_1, \cdots, x_j) - F_r{}^*(y_1, \cdots, y_j)$$

$$= -L^{R-1} T_{r0}\left(\sum_{l=1}^{j} \xi_l{}^2 + \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} \xi_{l_1}\xi_{l_2}\right) + o(1).$$

Also,

$$F_r(x_1, \cdots, x_j) = F_r{}^*(x_1, \cdots, x_j) + o(1)$$

from (2.15), (3.5) and (4.2), and Lemma 2 follows immediately.

We now turn to the proof of Lemma 3. It can be verified from (2.5), (2.6), (2.7) and (2.8) of [11] that, for $k+1 \leqq l \leqq j-1$ and all $n \geqq 0$,

$$\frac{\partial}{\partial z_l} I'_{kn}(z_{k+1}, \cdots, z_{j-1}) = -I'_{k,n+1}(z_{k+1}, \cdots, z_{j-1}) - I'_{k,n+1}(z_{k+1}, \cdots, z_{j-1}, z_l).$$

In particular, we have

$$y_l \frac{\partial}{\partial y_l} I(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$$
$$= -y_m^{-1}y_l\{I'_{01}(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$$
$$+ I'_{01}(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j, y_m^{-1}y_l)\}$$

for $l \neq m$, where $y_m = \max y_l$, and

$$y_m \frac{\partial}{\partial y_m} I(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$$
$$= \sum_{\substack{l=1 \\ l \neq m}}^{j} y_m^{-1}y_l\{I'_{01}(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$$
$$+ I'_{01}(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j, y_m^{-1}y_l)\}.$$

Also, from (3.4) of [11],

$$I(y_l^{-1}y_1, \cdots, y_l^{-1}y_{l-1}, y_l^{-1}y_{l+1}, \cdots, y_l^{-1}y_j)$$
$$= I(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$$
$$- L^{R-1}R_j(y_1, \cdots, y_j) \log y_l^{-1}y_m.$$

Therefore, if $\mu_{r1}, \cdots, \mu_{rk}$ are each $O(\mu_r^{*1+\epsilon_6})$ for some fixed positive number ϵ_6, but none of $\mu_{r,k+1}, \cdots, \mu_{rj}$ is $O(\mu_r^{*1+\epsilon_7})$ for any fixed positive number ϵ_7, we deduce from (2.4) of [11] and (2.19) that χ_{rl} can be written in the form

$$\chi_{rl} = \sum_{q_1, \cdots, q_j} d_{q_1, \cdots, q_j}{}^{(rl)} y_1^{q_1} \cdots y_j^{q_j}$$
$$+ y_m^j \sum_{q_1, \cdots, q_k} d'_{q_1, \cdots, q_k}{}^{(rl)} (y_m^{-1}y_1)^{q_1} \cdots (y_m^{-1}y_k)^{q_k} \tag{4.4}$$
$$+ O\{(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_8}\},$$

where the first summation on the right-hand side is taken over all non-negative integers $q_1, \cdots, q_j$ such that $\mu_{r1}^{q_1} \cdots \mu_{rj}^{q_j}$ is not $o\{(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_8}\}$ and the second summation is taken over all non-negative integers $q_1, \cdots, q_k$ such that $(\mu_r^{*-1}\mu_{r1})^{q_1} \cdots (\mu_r^{*-1}\mu_{rk})^{q_k}$ is not $o\{\mu_r^{*-j}(\mu_{r1} \cdots \mu_{rj})^{\frac{1}{2}+\epsilon_8}\}$. When $r = 3$ or 4, the coefficients d' are all zero since $\bar{\tau}_{r0}=0$ and the coefficients d are constants since $\tau_{rj_\nu}=0$ for $1\leqq\nu\leqq j$. When $r=1$ or 2, the coefficients d' are $O(1)$ and involve definite integrals of the type $I'_{kn}(y_m^{-1}y_{\nu_1},\cdots,y_m^{-1}y_{\nu_N})$, where $\nu_l\geqq k+1$ and $\nu_l\neq m$ for $1\leqq l\leqq N$, while the coefficients d are constants except when $q_1+\cdots+q_j=j$, in which case they are linear functions of the $\log y_l$ for $1\leqq l\leqq j$. Also, for all r and l, we have

$$d_{00\cdots0}{}^{(rl)}=1. \tag{4.5}$$

It can be seen from (4.4) that, in order to satisfy (3.3), it is necessary to choose the y_l to satisfy

$$\begin{aligned} n_ly_lL^R=T_{r0}\{&\sum_{q_1,\cdots,q_j} d_{q_1,\cdots,q_j}{}^{(rl)}y_1{}^{q_1}\cdots y_j{}^{q_j} \\ &+y_m{}^j\sum_{q_1,\cdots,q_k} d'_{q_1,\cdots,q_k}{}^{(rl)}(y_m^{-1}y_1)^{q_1}\cdots(y_m^{-1}y_k)^{q_k}\} \\ &+o\{(\mu_{r1}\cdots\mu_{rj})^{1+\epsilon_s}\}. \end{aligned} \tag{4.6}$$

We put

$$\begin{aligned} y_l=\mu_{rl}&\sum_{q_1,\cdots,q_j,q_1',\cdots,q_k'}\delta_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(rl)} \\ &\times\mu_{r1}{}^{q_1}\cdots\mu_{rj}{}^{q_j}(\mu_r{}^{*-1}\mu_{r1})^{q_1'}\cdots(\mu_r{}^{*-1}\mu_{rk})^{q_k'} \end{aligned} \tag{4.7}$$

for $1\leqq l\leqq j$, where the summation is taken over all non-negative $q_1,\cdots,q_j$ and over all non-negative $q_1',\cdots,q_k'$ such that

$$\mu_{r1}{}^{q_1}\cdots\mu_{rj}{}^{q_j}(\mu_r{}^{*-1}\mu_{r1})^{q_1'}\cdots(\mu_r{}^{*-1}\mu_{rk})^{q_k'}$$

is not $o\{(\mu_{r1}\cdots\mu_{rj})^{1+\epsilon_s}\}$. From (2.17),

$$n_l\mu_{rl}(\mu_{r1}\cdots\mu_{rj})=T_{r0}$$

and therefore, it is easily seen from (4.5) that, for all r and l,

$$\delta_{00\cdots0:00\cdots0}{}^{(rl)}=1.$$

Hence, the expressions for y_l given in (4.7) satisfy the equations (2.16). Next, by substituting for each y_l the finite series given in (4.7), it is not difficult to see that the same type of argument as that employed earlier in this section to show that $\Omega I(z_{m1},\cdots,z_{m,m-1},z_{m,m+1},\cdots,z_{mj})=o(1)$ can be applied here to show that every integral of the form $I'_{kn}(y_m^{-1}y_{\nu_1},\cdots,y_m^{-1}y_{\nu_N})$, where

$\nu_l \geqq k+1$ and $\nu_l \neq m$ for $1 \leqq l \leqq N$, can be approximated to the required accuracy by an expression of the form

$$\mu_r^{*N}(\mu_{r\nu_1}\cdots\mu_{r\nu_N})^{-1}\sum_{q_1,\cdots,q_j,q_1',\cdots,q_k'}\bar{d}_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(r)}$$
$$\times\,\mu_{r1}^{q_1}\cdots\mu_{rj}^{q_j}(\mu_r^{*-1}\mu_{r1})^{q_1'}\cdots(\mu_r^{*-1}\mu_{rk})^{q_k'}.$$

In this formula, the coefficients $\bar{d}$ are $O(1)$ and involve integrals of the type $I'_{kn}(\mu_r^{*-1}\mu_{r\nu_1},\cdots,\mu_r^{*-1}\mu_{r\nu_{N'}})$, where $\nu_l \geqq k+1$ and $\nu_l \neq m$ for $1 \leqq l \leqq N'$. Now, by substituting in (4.6) for each y_l the finite series given in (4.7) and by equating the coefficients of $\mu_{r1}^{q_1}\cdots\mu_{rj}^{q_j}(\mu_r^{*-1}\mu_{r1})^{q_1'}\cdots(\mu_r^{*-1}\mu_{rk})^{q_k'}$ on each side of the resulting equation, we can calculate the other coefficients δ successively. Thus, for example,

$$2\delta_{100\cdots0:00\cdots0}{}^{(rl)}+\sum_{\substack{\nu=1\\ \nu\neq l}}^{j}\delta_{100\cdots0:00\cdots0}{}^{(r\nu)}=d_{100\cdots0}{}^{(rl)}$$

for $1 \leqq l \leqq j$ and these equations can easily be solved to give each $\delta_{100\cdots0:00\cdots0}{}^{(rl)}$ in terms of the $d_{100\cdots0}{}^{(r\nu)}$ for $1 \leqq \nu \leqq j$. Also, for any $q_1,\cdots,q_j,q'_1,\cdots,q'_k$,

$$2\delta_{q_1\cdots q_j:q_1'\cdots q_k'}+\sum_{\substack{\nu=1\\ \nu\neq l}}^{j}\delta_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(r\nu)}=d_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{*(rl)},$$

where $d_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(rl)}$ involves the coefficients d, d' and $\bar{d}$ and the $\delta_{\nu_1\cdots\nu_j:\nu_1'\cdots\nu_k'}{}^{(r\nu)}$ for which every $\nu_l \leqq q_l$, every $\nu_l' \leqq q_l'$ and $0<\sum_{l=1}^{j}\nu_l+\sum_{l=1}^{k}\nu_l'<\sum_{l=1}^{j}q_l+\sum_{l=1}^{k}q_l'$. These equations can be solved to give each $\delta_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(rl)}$ in terms of the $d_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{*(r\nu)}$ for $1 \leqq \nu \leqq j$. Therefore, the coefficients $\delta_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(rl)}$ can be calculated successively, provided that we calculate first the $\delta_{q_1\cdots q_j:q_1'\cdots q_k'}{}^{(rl)}$ for which $\sum_{l=1}^{j}q_l+\sum_{l=1}^{k}q_l'=1$, then those for which $\sum_{l=1}^{j}q_l+\sum_{l=1}^{k}q_l'=2$, and so on. The y_l can therefore be expressed completely in terms of the μ_{rl} and this completes the proof of Lemma 3.

Next, we choose K' as the smallest integer for which $(\mu_r^{*-1}\mu_{rl})^{2K'} = o(\mu_{r1}^{*-j}\mu_{r1}\cdots\mu_{rj})$ for $1 \leqq l \leqq k$ and we deduce from (2.4) of [11] that

$$I(y_m^{-1}y_1,\cdots,y_m^{-1}y_{m-1},y_m^{-1}y_{m+1},\cdots,y_m^{-1}y_j)$$
$$=y_m^k(y_1,\cdots,y_k)^{-1}\sum_{n=0}^{2K'-1}I'_{kn}(y_m^{-1}y_{k+1},\cdots,y_m^{-1}y_{m-1},y_m^{-1}y_{m+1},\cdots,y_m^{-1}y_j)$$
$$\times R_n(y_m^{-1}y_1,\cdots,y_m^{-1}y_k)+o(1).$$

Then, by substituting for each y_l the finite series given in (4.7), the integral $I(y_m^{-1}y_1, \cdots, y_m^{-1}y_{m-1}, y_m^{-1}y_{m+1}, \cdots, y_m^{-1}y_j)$ can be expressed in terms of the μ_{rl} with an error of order $o(1)$. It follows that $\bar{\tau}_{r0}$ can be expressed in terms of the μ_{rl} with an error of order $o(1)$.

It is easily deduced from (2.17) and (3.1) that, since $\mu_{r1}, \cdots, \mu_{rk}$ are each $O(\mu_r^{*1+\epsilon_6})$ for some fixed positive number ϵ_6, $n_1, \cdots, n_k$ are each greater than $\bar{n}^{1+\epsilon_8}$ for some fixed positive number ϵ_8. Similarly, none of $n_{k+1}, \cdots, n_j$ is greater than $\bar{n}^{1+\epsilon_9}$ for any fixed positive number ϵ_9. For $r=1$ or 2, if $k=j-1$, or if $\mu_r^{*j}=o\{(\mu_{r1}\cdots\mu_{rj})^{\frac{1}{2}+\epsilon_8}\}$, i.e.

$$n_1\cdots n_j = o\left(\bar{n}^{j(j+1)/(j+\frac{1}{2}+\epsilon_8)}\right),$$

the y_l can be determined completely in terms of elementary functions. If, on the other hand, $k \leqq j-2$ and μ_r^{*j} is not $o\{(\mu_{r1}\cdots\mu_{rj})^{\frac{1}{2}+\epsilon_8}\}$, the y_l involve integrals of the type $I'_{kn}(\mu_r^{*-1}\mu_{r\nu_1}, \cdots, \mu_r^{*-1}\mu_{r\nu_{N'}}) = I'_{kn}(n_{\nu_1}^{-1}\bar{n}, \cdots, n_{\nu_{N'}}^{-1}\bar{n})$, where $k+1 \leqq \nu_l \leqq j$ and $\nu_l \neq m$ for $1 \leqq l \leqq N'$. We see from [11] that, when every z_l is real and rational, an exact formula can be obtained for both $I(z_1, \cdots, z_{j-1})$ and $I'_{kn}(z_{k+1}, \cdots, z_{j-1})$ and so, when the ratio of every pair of n_l is a fixed rational number for $k+1 \leqq l \leqq j$, the formula (3.2) of the theorem can be expressed entirely in terms of elementary functions of the n_l for $1 \leqq l \leqq j$. For $r=3$ or 4, since the d' are all zero, these difficulties do not arise and the formula (3.2) can always be expressed in terms of elementary functions of the n_l.

5. Proof of Lemma 4.

We denote the matrix of the real quadratic form

$$\sum_{l=1}^{j} u_l^2 + \sum_{l_1=1}^{j-1}\sum_{l_2=l_1+1}^{j} u_{l_1}u_{l_2}$$

by M_j, i.e. M_j is a square matrix of order j with every element on the principal diagonal equal to 1 and every other element equal to $\frac{1}{2}$. In order to show that the martix M_j is positive definite, it is sufficient to show that $\det M_n > 0$ for $1 \leqq n \leqq j$.

Now, for all $n > 1$,

$$\det M_n = \begin{vmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\cdots\frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2}\cdots\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{2}\cdots\frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\cdots 1 \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0\cdots 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{2}\cdots\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{2}\cdots\frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2}\cdots 1 \end{vmatrix}$$

$$= \tfrac{1}{2}(\det M_{n-1} + \det M'_{n-1}),$$

where $\det M'_n$ is the n-th order determinant which satisfies

$$\det M'_n = \begin{vmatrix} \frac12 & \frac12 & \frac12 & \frac12 & \cdots \frac12 \\ \frac12 & 1 & \frac12 & \frac12 & \cdots \frac12 \\ \frac12 & \frac12 & 1 & \frac12 & \cdots \frac12 \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \frac12 & \frac12 & \frac12 & \frac12 & \cdots 1 \end{vmatrix} = \begin{vmatrix} 0 & -\frac12 & 0 & 0 & \cdots 0 \\ \frac12 & 1 & \frac12 & \frac12 & \cdots \frac12 \\ \frac12 & \frac12 & 1 & \frac12 & \cdots \frac12 \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \frac12 & \frac12 & \frac12 & \frac12 & \cdots 1 \end{vmatrix}$$

$$= \tfrac12 \det M'_{n-1}.$$

Also, $M'_1 = \frac12$ and it follows by induction that $M'_n = 2^{-n}$. Therefore,

$$\det M_n = \tfrac12 \det M_{n-1} + 2^{-n}$$

and, since $\det M_1 = 1$, it can easily be proved, again by induction, that

$$(5.1) \qquad \det M_n = (n+1)2^{-n}$$

for all $n \geqq 1$. It follows immediately from (5.1) that M_j is positive definite.

It is well known that any real symmetric positive definite matrix can be expressed as the product of a non-singular real lower triangular matrix with its transpose. Therefore, there exists a square matrix W_j of order j, the elements of which are all real and satisfy $w_{kl} = 0$ for $k < l$, such that $M_j = W_j \tilde{W}_j$, where $\tilde{W}_j$ is the transpose of W_j. Then, since the determinant of W_j is equal to the product of its diagonal elements,

$$(5.2) \qquad \det M_j = w_{11}^2 \cdots w_{jj}^2.$$

We effect a change of the variables in the given integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\sum_{l=1}^{j} u_l^2 - \sum_{l_1=1}^{j-1} \sum_{l_2=l_1+1}^{j} u_{l_1} u_{l_2}\right) du_1 \cdots du_j$$

by means of the transformation $v_l = \sum_{\nu=l}^{j} w_{\nu l} u_l$ for $1 \leqq l \leqq j$. It is easily seen from (5.1) and (5.2) that the Jacobian of the above transformation is numerically equal to $(j+1)^{\frac12} 2^{-\frac12 j}$ and therefore, the integral is equal to

$$(j+1)^{-\frac12} 2^{\frac12 j} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\sum_{l=1}^{j} v_l^2\right) dv_1 \cdots dv_j$$

$$= (j+1)^{-\frac12} 2^{\frac12} \prod_{l=1}^{j} \left\{ \int_{-\infty}^{\infty} \exp(-v_l^2)\, dv_l \right\}$$

$$= (j+1)^{-\frac12} (2\pi)^{\frac12 j}.$$

THE UNIVERSITY, ABERDEEN, SCOTLAND.

REFERENCES.

[1] G. H. Hardy and S. Ramanujan, *Proceedings of the London Mathematical Society*, ser. (2), vol. 17 (1918), pp. 75-115.

[2] F. C. Auluck, *Proceedings of the Cambridge Philosophical Society*, vol. 49 (1953), pp. 72-83.

[3] G. Meinardus, *Mathematische Annalen*, vol. 132 (1956), pp. 333-346.

[4] V. S. Nanda, *Proceedings of the Cambridge Philosophical Society*, vol. 53 (1957), pp. 272-277.

[5] E. M. Wright, *Proceedings of the American Mathematical Society*, vol. 7 (1956), pp. 880-890.

[6] ———, *Proceedings of the London Mathematical Society*, ser. (3), vol. 7 (1957), pp. 150-160.

[7] ———, *American Journal of Mathematics*, vol. 80 (1958), pp. 643-658.

[8] ———, *Quarterly Journal of Mathematics*, ser. (2), vol. 10 (1959), pp. 60-69.

[9] ———, *Proceedings of the London Mathematical Society*, ser. (3), vol. 8, pp. 312-320.

[10] M. M. Robertson, *Proceedings of the Edinburgh Mathematical Society*, vol. 12 (1960), pp. 31-40.

[11] ———, *Proceedings of the Royal Society of Edinburgh*, vol. LXV, A (1961), pp. 283-309.

HOMOLOGY OF ITERATED LOOP SPACES.*

By Eldon Dyer[1] and R. K. Lashof.[1]

One of the impotrant problems of topology is to determine the stable homotopy groups $S_n(X)$ of a space X; i.e., $\varinjlim \pi_{i+n}(s^iX)$, where s^iX denotes the i-th reduced suspension of X and $\pi_{i+n}(s^iX) \to \pi_{i+1+n}(s^{i+1}X)$ is the Freudenthal Suspension Homomorphism. The Freudenthal Suspension Theorem asserts that this homomorphism is an isomorphism for $n < i-1$; and so, the group $S_n(X)$ is isomorphic to $\pi_{i+n}(s^iX) = \pi_n(\Omega^i s^iX)$ for $n < i-1$, where Ω^i denotes the i-th iterated loop space. As there is a natural imbedding of $\Omega^i s^iX$ in $\Omega^{i+1}s^{i+1}X$, it is clear that if $Q(X)$ denotes the direct limit of the spaces $\Omega^i s^iX$ under this imbedding, then

$$S_n(X) = \pi_n(Q(X)).$$

The space $Q(X)$ is in a well-defined sense an "infinite loop space." Araki and Kudo found [1] that iterated loop spaces have additional structure beyond the H-space structure to be found in all loop spaces. This structure affords a measure of the lack of commutativity of the H-space multiplication. In terms of this structure they defined mod 2 "homology operations" much analogous to the Steenrod Squares and were able to compute in terms of these operations the mod 2 homology structure of $\Omega^i S_n$, $i < n$, where S_n denotes the n-sphere.

Difficulties arise in attempting to extend their definitions to the mod p case, for p an odd prime. Browder in his thesis [3] defined a type of mod p homology operation and computed the mod p homology of $\Omega^n s^nX$, where X is arcwise connected and $p > n/2$.

In this paper the structure on iterated loop spaces found by Araki and Kudo is further developed into a form which appears dual to the diagonal structure used by Steenrod in defining cohomology reduced powers. In terms of this structure operations on homology classes are defined, *homology extended powers.* These operation have properties analogous to those of the Steenrod operations. Also these operations are in a certain sense "stable" homology operations, as distinct from Browder's.

---

* Received January 3, 1961; revised December 11, 1961.

[1] This research was supported in part by the National Science Foundation under contract number G-10369.

As applications, we compute the mod p homology of $\Omega^i S_n$, $i < n$, and $Q(sX)$, for X an arcwise connected space. Also, partial results on $\Omega^i s^i X$ are obtained. Unlike the mod 2 case, $H_*(\Omega^i s^n X; Z_p)$ is not generated by homology operations (as defined herein) on $\operatorname{im} H_*(s^{n-i}X; Z_p)$. However, $Q(sX)$ is, roughly speaking, the smallest space containing sX for which the homology operations are all defined, and, in fact, the homology of $Q(sX)$ is generated by homology operations on the homology classes of sX. This may be compared with the result of James [8] that ΩsX is in some sense the smallest H-space containing X. Also, we show that in a space $\Omega^i s^n X$, $i < n$, a number of cohomology operations are trivial.

I. H^n-spaces.

1. *Definition of H^n-spaces.* Let Σ_p denote the group of permutations on p symbols and let $J^n\Sigma_p$ denote the n-th-join of Σ_p with itself (in the sense of Milnor [9]). Briefly, a point of $J^n\Sigma_p$ is determined by a sequence $t_0, \cdots, t_{n-1}$ of real numbers such that $t_i \geqq 0$ and $t_0 + \cdots + t_{n-1} = 1$, and an element $\sigma_i \in \Sigma_p$ for each i such that $t_i \neq 0$. Such a point is denoted by $(t_0\sigma_0 \oplus \cdots \oplus t_{n-1}\sigma_{n-1})$. The set of these points is given the strong topology.

A space X is an $H^n{}_p$-space, $n \geqq 0$, provided it has an associative multiplication with unit e and there is a map (where X^p is the Cartesian product of X with itself p times)

$$\theta^n{}_p \colon J^{n+1}\Sigma_p \times X^p \to X$$

which is

1.1. *Σ_p-equivariant*; i.e., for each $\sigma \in \Sigma_p$

$$\theta^n{}_p(t_0\sigma_0 \oplus \cdots \oplus t_n\sigma_n; x_1, \cdots, x_p) = \theta^n{}_p(t_0\sigma_0\sigma^{-1} \oplus \cdots \oplus t_n\sigma_n\sigma^{-1}; x_{\sigma(1)}, \cdots, x_{\sigma(p)})$$

and

1.2. *normalized*; i. e., for each $\sigma \in \Sigma_p$

$$\theta^n{}_p(0 \oplus \cdots \oplus 0 \oplus 1\cdot\sigma; x_1, \cdots, x_p) = x_{\sigma(1)} \cdot \cdots \cdot x_{\sigma(p)}.$$

Let $J^\infty\Sigma_p = \varinjlim J^n\Sigma_p$, where $J^n\Sigma^p \subset J^{n+1}\Sigma_p$ is given by

$$(t_1\sigma_1 \oplus \cdots \oplus t_n\sigma_n) \to (0 \oplus t_1\sigma_1 \oplus \cdots \oplus t_n\sigma_n).$$

A space X is an $H^\infty{}_p$-space if there is a map $\theta^\infty{}_p \colon J^\infty\Sigma_p \times X^p \to X$ such that for each n, $\theta^\infty{}_p \mid J^{n+1}\Sigma_p \times X^p$ makes X an $H^n{}_p$-space.

The H_n-spaces defined by Araki and Kudo [1] are $H^n{}_2$-spaces in the above sense. It is also clear that an H^{n+1}-space is *a fortiori* an H^n_p-space. Araki and Kudo showed that an $(n+1)$-st loop space is an H^n_2-space. The corresponding theorem is true here also.

A space X is a special H^n_p-space if it is an H^n_p-space and

1.3. the map θ^n_p is *projective*; i.e.,

$$\theta^n_p(w; e, \cdots, e, \overset{i}{x}, e, \cdots, e) = x$$

for all i, $1 \leqq i \leqq p$, $w \in J^{n+1}\Sigma_p$ and $x \in X$.

THEOREM 1.1. *An H-space* (in the usual sense = associative multiplication with unit) *is a special H^0_p-space for all p. If X is a special H^n_p-space, then ΩX, the loop space of X, is a special H^{n+1}_p-space.*

In particular, we note that *an* $(n+1)$-st-*loop space is an H^n_p-space for all p.*

Proof. We use the "Moore loop space." This is of the same homotopy type as the ordinary loop space and is more convenient for our purposes. Recall that a point of $\Omega(X, e)$ is a pair (w, r), $r \geqq 0$ a real number and w a map: $w\colon [0, r] \to X$, such that $w(0) = w(r) = e$, and if (w, r) and (w', r') are points of $\Omega(X, e)$, then

$$(w, r) \cdot (w', r') = (w'', r'')$$

is defined by

$$\begin{aligned} w''(t) &= w(t), 0 \leqq t \leqq r \\ &\qquad w'(t-r), r \leqq t \leqq r + r'. \end{aligned}$$

For X a special H^n_p-space, we define the loop $(t_i \geqq 0, \sum t_i = 1, 0 \leqq s \leqq 1)$

1.4. $\theta^{n+1}_p((1-s)t_0\sigma_0 \oplus \cdots \oplus (1-s)t_n\sigma_n \oplus s\epsilon;\ (w_1, r_1), \cdots (w_p, r_p))(t)$
for $0 \leqq t \leqq r = r_1 + \cdots + r_p$, in the space ΩX to be

$$\theta^n_p(t_0\sigma_0 \oplus \cdots \oplus t_n\sigma_n; w^s_1(t), \cdots, w^s_p(t)),$$

where

$$\begin{aligned} w^s_i(t) &= \bar{w}_i(t - s(r_1 + \cdots + r_{i-1})) \text{ and} \\ \bar{w}_i(t) &= w_i(0), t < 0 \\ &\qquad w_i(t), 0 \leqq t \leqq r_i \\ &\qquad w_i(r_i), r_i < t. \end{aligned}$$

This defines θ^{n+1}_p for $\sigma_{n+1} = \epsilon$, the unit element of Σ_p; for σ_{n+1} different from ϵ, we define θ^{n+1}_p by equivariance.

It follows immediately from this definition that $\theta^{n+1}{}_p$ satisfies conditions 1.1, 1.2 and 1.3.

A map $f\colon X \to \bar{X}$ of $H^n{}_p$-spaces will be called an $H^n{}_p$-*map* if

$$\begin{tikzcd}
J^{n+1}\Sigma_p \times X^p \arrow[r, "\theta^n{}_p"] \arrow[d, "1 \times f^p"] & X \arrow[d, "f"] \\
J^{n+1}\Sigma_p \times \bar{X}^p \arrow[r, "\bar{\theta}^n{}_p"] & \bar{X},
\end{tikzcd}$$

commutes up to a Σ_p-equivariant homotopy. In particular, if $X = \Omega^{n+1}Y$, $\bar{X} = \Omega^{n+1}\bar{Y}$ and f is the map induced by a map $g\colon Y \to \bar{Y}$, then f is an $H^n{}_p$-map.

Iterated loop spaces have additional structure beyond that of being $H^n{}_p$-spaces. This structure is necessary to establish "Adem Relations" for the homology operations.[2] We now describe it.

Recall that the *wreath product* $\Sigma_m \int \Sigma_l$ of Σ_m by Σ_l is the subgroup of Σ_{ml} described as follows: consider Σ_{ml} as the permutation group on the ml objects $x_1, \cdots, x_{ml}$. $\Sigma_m \int \Sigma_l$ is the subgroup of Σ_{ml} generated by $(\Sigma_m)^l$ and Σ_l, where the i-th factor of $(\Sigma_m)^l$ acts on the i-th block of m objects (leaving the others fixed), $i = 1, \cdots, l$, and Σ_l acts by permuting the l such blocks. Explicitly, if $\sigma \in \Sigma_l$ and $(\tau_1, \cdots, \tau_l) \in (\Sigma_m)^l$, then $\sigma[\tau_1, \cdots, \tau_l]$ denotes an element of $\Sigma_m \int \Sigma_l$ and multiplication is given by

$$\sigma[\tau_1, \cdots, \tau_l] \cdot \bar{\sigma}[\bar{\tau}_1, \cdots, \bar{\tau}_l] = \sigma\bar{\sigma}[\tau_{\bar{\sigma}(1)}\bar{\tau}_1, \cdots, \tau_{\bar{\sigma}(l)}\bar{\tau}_l].$$

It will be convenient for our purposes to have an explicit description of the simplicial structures of the complexes $J^r\Sigma_l$ and $J^s\Sigma_m \times J^t\Sigma_n$. A *vertex* v of $J^r\Sigma_l$ is a point of the form

$$v = v(i) = (0 \oplus \cdots \oplus 0 \oplus \overset{i}{\sigma} \oplus 0 \oplus \cdots \oplus 0),$$

$0 \leqq i \leqq r-1$. We call i the *weight* of v, and write $w(v) = i$. *A q-simplex* in $J^r\Sigma_l$ is of the form $(\tau_0(i_0), \cdots, \tau_q(i_q))$, $\tau_i \in \Sigma_l$, $i_0 < \cdots < i_q < r$, and consists of all points $(a_0\sigma_0 \oplus \cdots \oplus a_{r-1}\sigma_{r-1})$ in $J^r\Sigma_l$ whose coordinates are all zero except in the i_j places and $\sigma_{i_j} = \tau_j$, $j = 0, \cdots, q$.

For u a vertex of $J^s\Sigma_m$ and v a vertex of $J^t\Sigma_n$, (uv) will denote the corresponding vertex of $J^s\Sigma_m \times J^t\Sigma_n$. Let $w(uv) = w(u) + w(v)$. Given p- and q-simplices of $J^s\Sigma_m$ and $J^t\Sigma_n$, respectively, we simplicially subdivide their direct product in the usual way; i.e., an r-simplex of their product is a sequence $(u_0v_0,\cdots,u_rv_r)$, where u_i and v_i are vertices of the p and q-simplices, respectively, satisfying

---

[2] The Adem relations are not explicitly used in the examples we compute in this paper, so that the remainder of this section may be omitted on first reading.

(i) $w(u_iv_i) < w(u_{i+1}v_{i+1})$ and

(ii) $w(u_i) \leqq w(u_{i+1})$ and $w(v_i) \leqq w(v_{i+1})$.

(Note that $w(v_i) = w(v_{i+1})$ implies $v_i = v_{i+1}$.)

$J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l$ is a $\Sigma_m \int \Sigma_l$-free complex; i.e. the i-th factor of $(\Sigma_m)^l$ acts on the i-th factor of $(J^{r+1}\Sigma_m)^l$ and leaves the other factors as well as $J^{s+1}\Sigma_l$ fixed and Σ_l acts by permuting the factors in $(J^{r+1}\Sigma_m)^l$ and acts on $J^{s+1}\Sigma_l$ in the usual way.

Corresponding to the inclusion map of $\Sigma_m \int \Sigma_l$ in Σ_{ml}, there is a $\Sigma_m \int \Sigma_l$-equivariant map

$$\psi\colon J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l \to J^{t+1}\Sigma_{ml},\quad t = s + rl,$$

which we describe explicitly. A vertex in $J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l$ is of the form

$$v = \sigma(i)\tau_1(j_1)\cdots\tau_l(j_l),\ w(v) = i + j_1 + \cdots + j_l = k.$$

Let $\psi(v) = [\tau_1,\cdots,\tau_l](k)$. The condition (i) above implies that ψ is simplicial. Equivariance is immediate from the definitions.

An H^t-*space* is an H-space X which is an $H^t{}_p$-space for every positive integer p and if $t = s + rl$, then

$$\begin{tikzcd}
J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m \times X^m)^l \arrow[r, "i \times (\theta^r{}_m)^l"] \arrow[d, "\bar\psi"] & J^{s+1}\Sigma_l \times X^l \arrow[d, "\theta^s{}_l"] \\
J^{t+1}\Sigma_{ml} \times X^{ml} \arrow[r, "\theta^t{}_{ml}"] & X
\end{tikzcd}$$

commutes up to a $\Sigma_m \int \Sigma_l$-equivariant homotopy, where $\bar\psi$ is the composition

$$J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m \times X^m)^l \to J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l \times X^{ml}$$
$$\to J^{t+1}\Sigma_{ml} \times X^{ml}.$$

THEOREM 1.2. *Every $(t+1)$-st loop space is an H^t-space.*

Proof. Let $t = s + rl$. We first define a map

$$\phi\colon J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l \to J^{t+2}\Sigma_l \times (J^{t+2}\Sigma_m)^l$$

which is $\Sigma_m \int \Sigma_l$-equivariantly homotopic to the standard inclusion of $J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l$ in $J^{t+2}\Sigma_l \times (J^{t+2}\Sigma_m)^l$. For each vertex $v = \sigma(t)\tau_1(j_1)\cdots\tau_l(j_l)$, let $\phi(v) = \sigma(k)\tau_1(k)\cdots\tau_l(k)$, where $k = i + j_1 + \cdots + j_l$.

This map is simplicial, as above, and since $\dim(J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l) = t$ and $J^{t+2}G$, G any finite group, is $(t+1)$-connected, any two $\Sigma_m \int \Sigma_l$-equivariant maps of $J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m)^l$ into $J^{t+2}\Sigma_l \times (J^{t+2}\Sigma_m)^l$ are $\Sigma_m \int \Sigma_l$-equivariantly homotopic.

It is this sufficient to show that the diagram

$$\begin{tikzcd}
J^{s+1}\Sigma_l \times (J^{r+1}\Sigma_m \times X^m)^l \arrow[r, "\bar{\phi}"] \arrow[dd, "\bar{\psi}"] & J^{t+2}\Sigma_l \times (J^{t+2}\Sigma_m \times X^m)^l \arrow[d, "i \times (\theta)^l"] \\
& J^{t+2}\Sigma_l \times X^l \arrow[d, "\theta"] \\
J^{t+1}\Sigma_{ml} \times X^{ml} \arrow[r, "\theta"] & X
\end{tikzcd}$$

is commutative. We will show this by an induction on dimension of cells of $J^{t+1}\Sigma_{ml}$ in the image of ψ.

Since

$$\theta^r{}_n(s_0\sigma_0 \oplus \cdots \oplus s_{r-2}\sigma_{r-2} \oplus 0; \lambda_1, \cdots, \lambda_n)(t)$$
$$= \theta^{r-1}{}_n(s_0\sigma_0 \oplus \cdots \oplus s_{r-2}\sigma_{r-2}; \lambda_1(t), \cdots, \lambda_n(t)),$$

the result is immediate for 0-cells. Assume then that the diagram commutes for p-simplices in the image of ψ; i. e., a simplex

$$(\sigma^0[\tau^0{}_1, \cdots, \tau^0{}_l](k^0), \cdots, \sigma^p[\tau^p{}_1, \cdots, \tau^p{}_l](k^p)),$$

where

$$\sigma^a[\tau^a{}_1, \cdots, \tau^a{}_l](k^a) = \psi(\sigma^a(i^a)\tau^a{}_1(j^a{}_1) \cdots \tau^a{}_l(j^a{}_l))$$

and $k^a = i^a + j^a + \cdots + j^a{}_l$. Thus, we assume

$$\theta^t{}_{ml}(\bigoplus_{a=0}^{p} s_a\sigma^a[\tau^a{}_1, \cdots, \tau^a{}_l](k^a); \lambda_1, \cdots, \lambda_{ml})$$
$$= \theta^{t+1}{}_l(\bigoplus_{a=0}^{p} s_a\sigma^a(k^a); \theta^{t+1}{}_m(\bigoplus_{a=0}^{p} s_a\tau^a{}_1(k^a); \lambda_1, \cdots, \lambda_m), \cdots,$$
$$\theta^{t+1}{}_m(\bigoplus_{a=0}^{p} s_a\tau a_l(k^a); \lambda_{m(l-1)+1}, \cdots, \lambda_{ml})).$$

We shall use the formula (1.4) to reduce the case of the $(p+1)$-simplex to that of the p-simplex. It is sufficient to prove the formula when the $(p+1)$-st vertex is the unit element of $\Sigma_m \int \Sigma_l$ as the action of the group on the two sides of the equation corresponds.

$$
\begin{aligned}
&\theta^{t}{}_{ml}\Big(\bigoplus_{a=0}^{p} s_a(1-s)\sigma^a[\tau^a{}_1,\cdots,\tau^a{}_l]\,(k^a)\oplus s\epsilon(k^{p+1})\,;\lambda_1,\cdots,\lambda_{ml}\Big)\,(t)\\
&=\theta^{t-1}{}_{ml}\Big(\bigoplus_{a=0}^{p} s_a\sigma^a[\tau^a{}_1,\cdots,\tau^a{}_l](k^a)\,;\lambda^s{}_1(t),\cdots,\lambda^s{}_{ml}(t)\Big)\\
&=\theta^{t}{}_{l}\Big(\bigoplus_{a=0}^{p} s_a\sigma^a(k^a)\,;\theta^{t}{}_{m}\Big(\bigoplus_{a=0}^{p} s_a\tau^a{}_1(k^a)\,;\lambda_1(t),\cdots\lambda_m\Big(t-s\sum_{i=1}^{m-1}u_i\Big)\Big),\cdots,\\
&\qquad\qquad\theta^{t}{}_{m}\Big(\bigoplus_{a=0}^{p} s_a\tau^a{}_l(k^a)\,;\lambda_{m(l-1)+1}\Big(t-s\sum_{i=1}^{m(l-1)}u_i\Big),\cdots,\\
&\qquad\qquad\qquad\lambda_{ml}\Big(t-s\sum_{i=1}^{ml-1}u_1\Big)\Big)\Big)\\
&=\theta^{t}{}_{l}\Big(\bigoplus_{a=0}^{p} s_a\sigma^a(k^a)\,;\theta^{t+1}{}_{m}\Big(\bigoplus_{a=0}^{p} s_a(1-s)\tau^a{}_1(k^a)\ \oplus\ s\epsilon(k^{p+1})\,;\\
&\qquad\qquad\qquad\lambda_1,\cdots,\lambda_m\Big)(t),\cdots,\\
&\quad\theta^{t+1}{}_{m}\Big(\bigoplus_{a=0}^{p} s_a(1-s)\tau^a{}_l(k^a)\oplus s\epsilon(k^{p+1})\,;\lambda_{m(l-1)+1},\cdots,\lambda_{ml}\Big)\Big(t-s\sum_{i=1}^{m(l-1)}u_i\Big)\Big)\\
&=\theta^{t+1}{}_{l}\Big(\bigoplus_{a=0}^{p} s_a(1-s)\sigma^a(k^a)\oplus s\epsilon(k^{p+1})\,;\theta^{t+1}{}_{m}\Big(\bigoplus_{s=0}^{p} s_a(1-s)\tau^a{}_1(k^a)\oplus s\epsilon(k^{p+1})\,;\\
&\qquad\qquad\qquad\lambda_1,\cdots,\lambda_m\Big),\cdots,\\
&\qquad\qquad\theta^{t+1}{}_{m}\Big(\bigoplus_{s=0}^{p} s_a(1-s)\tau^a{}_l(k^a)\oplus s\epsilon(k^{p+1})\,;\lambda_{m(l-1)+1},\cdots,\lambda_{ml}\Big)\Big)(t).\\
&=\theta^{t+1}{}_{l}(\iota\times(\theta^{t+1}{}_{m})^{l})\bar{\phi}\Big(\bigoplus_{s=0}^{p} s_a(1-s)\sigma^a(i^a)\tau^a{}_1(j^a{}_1)\cdots\tau^a{}_l(j^a{}_l)\\
&\qquad\qquad\oplus s\epsilon(i^{p+1})\epsilon(j^{p+1}{}_1)\cdots\epsilon(j^{p+1}{}_l)\,;\lambda_1,\cdots,\lambda_{ml}\Big)(t).
\end{aligned}
$$

Thus, $\theta^{t}{}_{ml}\circ\bar{\psi}=\theta^{t+1}{}_{l}\circ(\iota\times(\theta^{t+1}{}_{m})^{l})$ on $(p+1)$-simplices of $J^{t+1}\Sigma_{ml}$ in the image of ψ. The induction is complete.

II. Definition and first properties of homology operations. Our definitions are closely patterned on those of Steenrod ([15], [16]) for cohomology operations. The homology operations will have many properties similar to those of cohomology operations.

For any space A, let $C(A)$ denote the normalized cubical chains of A. An action of the symmetric group Σ_p on A induces an action of Σ_p *on* $C(A)$. Now let X be an $H^n{}_p$-space and $\theta=\theta^n{}_p\colon J^{n+1}\Sigma_p\times X^p\to X$ be the Σ_p-equivariant map defined in Section 1. The induced chain map $\theta\colon C(J^{n+2}\Sigma_p\times X^p)\to C(X)$ is then an equivariant chain map. Further, let $h\colon C(J^{n+1}\Sigma_p)\oplus C(X^p)\to C(J^{n+1}\Sigma_p\times X^p)$, $h(w\otimes a)=w^*a$, $w\in C(J^{n+1}\Sigma_p)$, $a\in C(X^p)$, be the standard chain equivalence (Eilenberg-Zilber Theorem). Then

$$\theta\circ h\colon C(J^{n+1}\Sigma_p)\otimes C(X^p)\to C(X)$$

is again clearly equivariant. If $\pi\subset\Sigma_p$, then $J^{n+1}\pi\subset J^{n+1}\Sigma_p$ and every Σ_p-equivariant map is π-equivariant. Summarizing we have:

LEMMA 2.1. *For any subgroup* π *of* Σ_p, *the equivariant map* $\theta\colon J^{n+1}\pi \times X^p \to X$ *induces an equivariant chain group* $\theta \circ h\colon C(J^{n+1}\pi) \otimes C(X^p) \to C(X)$ *on normalized cubical chains, where* $h(w \otimes a) = w^* a$, $w \in C(J^{n+1}\pi)$, $a \in C(X^p)$.

Let $C(X)^p = C(X) \otimes \cdots \otimes C(X)$, p times, and let Σ_p act on $C(X)^p$ in the usual way by permitting factors (with the appropriate sign [13]). In Lemma 2.1, we would like to substitute $C(X)^p$ for $C(X^p)$ as it is easier to handle. Although one may write down an explicit equivariant chain map from $C(X)^p$ to $C(X^p)$, we would like to know how unique the result is and whether we are losing any information by this substitution. The answer is given by the following theorem and corollary

Let K and L be covariant functors on a category $\mathcal{A}$ with values in the category $\partial\mathcal{G}$ of chain complexes and let $f\colon K \to L$ be any map. Let π be a finite group, and assume that K and L are π-functors; i.e., π is a group of natural transformations. If W and V are any π-chain complexes (assumed free over the integers) we make $W \otimes K$ and $V \otimes L$ into π.functors by having π act on both factors. For a complex W we will write $W^{(n)}$ for the n-skeleton: i.e. $W^{(n)} = \sum_{i=0}^{n} W_i$. (See [5] for notation.)

THEOREM 2.1. *If* W *is* π*-free, if* K *is representable and* L *is acyclic (for some set of models* $\mathcal{M} \subset \mathcal{A}$, *and if* f *is equivariant in dimension zero; then given any* π*-equivariant chain map* $t\colon W \to V$.

(a) $\exists$ *a* π*-equivariant map* $F\colon W \otimes K \to V \otimes L$, *satisfying*

(1) $F(W^{(n)} \otimes K(X)) \subset V^{(n)} \otimes L(X)$, *all* n,

(2) $F(w \otimes a) = t(w) \otimes f(a)$, $w \in W$, $a \in K_0(X)$.

(b) *If* t, $t^1\colon W \to V$ *are* π*-equivariantly chain homotopic, and* F, F^1 *are any two maps satisfying* (1) *and* (2) *above for* t, t^1 *respectively, then* F *and* F^1 *are* π*-equivariantly chain homotopic.*

(c) *We may further choose* F *so that given any two dimensional* π*-generator* e_0 *of* W, $F(e_0 \otimes a) = t(e_0) \otimes f(a)$, $a \in K(X)$.

Proof. Choose a set S of π-free generators of W. Define $F(e_0 \otimes a) = t(e_0) \otimes f,a)$, $a \in K(X)$, all o-dim π generators $e_0 \in S$; and extend to all of $W_0 \otimes K(X)$ by equivariance. Since f is a natural chain map, F will be a natural chain map, equivariant by definition, and clearly satisfying (1) and (2).

Assume F has been defined on $W^{(i)} \otimes K(X)$, then define it on $W_{i+1} \otimes K_0(X)$

by (2). Now assume F is defined on $W^{(i+1)} \otimes K(X)^{(r)}$; we show that F may be extended to $W^{(i+1)} \otimes K(X)^{(r+1)}$. Thus F will be extended to all of $W \otimes K(X)$ by induction.

For any $(i+1)$ dim π-generator $e_{i+1} \in S$, and any $m_{r+1} \in K_{r+1}(M)$, $M \subset \mathfrak{M}$, consider $e_{i+1} \otimes m_{r+1}$. Then

$$F(\partial(e_{i+1} \otimes m_{r+1})) = F(\partial e_{i+1} \otimes m_{r+1}) + (-1)^{i+1} F(e_{i+1} \otimes \partial m_{r+1})$$

is defined by the induction assumption, and is a cycle. Further, it is of the form $\sum_\rho v^\rho \otimes b^\rho$, $\dim v \leqq i+1$, $v \in V$, $b^\rho \in L(M)$. Let $\{v^\lambda{}_{i+1}\}$, be a set of generators of V_{i+1} over the integers. Then $\sum_\rho v^\rho \otimes b^\rho = \sum_\lambda v^\lambda{}_{i+1} \otimes b^\lambda + \sum_\sigma v^\sigma \otimes b^\sigma$, $\dim v^\sigma \leqq i$. Since $\sum_\rho v^\rho \otimes b^\rho$ is a cycle, and

$$\partial(v^\lambda{}_{i+1} \otimes b^\lambda) = \partial v_{i+1} \otimes b^\lambda + (-1)^{i+1} v^\lambda{}_{i+1} \otimes \partial b^\lambda;$$

it follows (since the $v^\lambda{}_{i+1}$ are independent) that $v^\lambda{}_{i+1} \otimes \partial b^\lambda = 0$ and hence $\partial b^\lambda = 0$. But $b^\lambda \in L_r(M)$, and $H_r(L(M)) = 0$ if $r > 0$, so that $b^\lambda = \partial c^\lambda$, $c^\lambda \in L_{r+1}(M)$. On the other hand, if $r = 0$,

$$F(e_{i+1} \otimes \partial m_{r+1}) = t(e_{i+1}) \otimes f(\partial m_{r+1}) = t(e_{i+1}) \otimes \partial f(m_{r+1})$$

by (2), and the form is the same as above, taking $c^\lambda = f(m_{r+1})$.

Consider now $F\partial(e_{i+1} \otimes m_{r+1}) - \partial(\sum v^\lambda{}_{i+1} \otimes e^\lambda)$. This is again a cycle and is of the form $\sum v^\rho \otimes b^\rho$, $\dim v^\rho \leqq t$. Proceeding as above, we got finally:

$$F(\partial(e_{i+1} \otimes m_{r+1})) = \partial(\sum_\tau v^\tau \otimes c^\tau), \dim v^\tau \leqq i+1.$$

Hence we may set $F(e_{i+1} \otimes m_{r+1}) = \sum_\tau v \otimes c^\tau$. Since K is representable, and hence any $x_{r+1} \in K_{r+1}(X)$ is of the form $K(\phi)(m_{r+1}) = x_{r+1}$, where $\phi\colon M \to X$ and (ϕ, m_{r+1}) is the chosen representative; we can define

$$F(e_{i+1} \otimes x_{r+1}) = \otimes L(\phi)(F(e_{i+1} \otimes m_{r+1})),$$

$i\colon V \to V$ the identity. F is then extended to all of $W^{(i+1)} \otimes K(X)^{(r+1)}$ by equivariance. It is easy to check that F has all the desired properties, including (c). This completes the induction argument.

The proof of (b) is entirely analogous, using condition (2) and the fact that t and t^1 are π equivariantly chain homotopic, to get the induction started. We leave the details to the reader.

COROLLARY. *Let W be any π-free chain complex, $\pi \subset \Sigma_p$; then*

(a) *Given any π-equivariant chain map $t\colon W \to C(J^{n+1}\pi)$, there exist a natural π-equivariant chain map $F\colon W \otimes C(X)^p \to C(J^{n+1}\pi) \otimes C(X^p)$ such that*

1) $F(w \otimes x_1 \otimes \cdots \otimes x_p) = t(w) \otimes x_1 * \cdots * x_p$, $\dim x_j = 0$, $j = 1, \cdots, p$

2) $F(W^{(i)} \otimes C(X)^p) \subset C(J^{n+1}\pi)^{(i)} \otimes C(X^p)$

(b) *If* $W^{(i)} = 0$ *for* $i > n$ (*e.g., if* W *is the* n*-skeleton of any* π*-free complex*), *then a map* t: $W \to C(J^{n+1}\pi)$ *always exists, and any two maps* F, F^1 *with equivariant chain maps* t, t^1 *resp., satisfying* 1), 2), *of* (a), *are equivariantly chain homotopic when restricted to* $W^{(n-1)} \otimes C(X)^p$.

If further, $H_0(W) = Z$, $H_i(W) = 0$, $0 < i < n$, *then* $\exists$ *a map* G: $C(J^{n+1}\pi)^{(n-1)} \otimes C(X^p) \to W \otimes C(X)^p$, *so that* GF: $W^{(n-1)} \otimes C(X)^p \to W \otimes C(X)^p$ *and* FG: $C(J^{r+1}\pi)^{(n-1)} \otimes C(X^p) \to C(J^{n+1}\pi) \otimes C(X^p)$ *are equivariantly chain homotopic to the respective inclusion maps.*

(c) *Any two maps* F, F^1 *for equivariantly chain homotopic maps* t, t^1, *satisfying* 1), 2) *of* (a), *are equivariantly chain homotopic; if* t *is an equivariant chain equivalence,* F *is an equivariant chain equivalence. In particular if* $W = C(J^{n+1}\pi)$ *and* t *is the identity, then any* F *satisfying* 1) *and* 2) *of* (a) *is an equivarient chain equivalence.*

(d) *If* e_0 *is any zero dimensional* π*-generator of* W, *then we may further choose* F *so that* $F(e_0 \otimes x_1 \cdots \otimes x_p) = t(e_0) \otimes x_1 * \cdots * x_p$, $x_i \in C(X)$. *Also,* t *may be chosen so that* $t(e_0)$ *is any given zero cell in* $C(J^{n-1}\pi)$.

From the comparison theorem applied to the Cartan-Leray spectral sequence for covering spaces, one gets immediately the following two Lemmas:

LEMMA. *If* W *is a* π*-free complex,* μ: $C \to D$ *a map of* π*-complexes such that* μ_*: $H(C) \cong H(D)$, *then* (π *acting trivially on* G):

$$\mu_{\pi_*}\colon H(W \otimes_\pi C; G) = H(W \otimes_\pi D; G)$$

LEMMA. *If* W *is a* π*-free complex,* C *is any* π*-complex,* G *is a field of coefficients or the cycles of* $C \otimes G$ *are a direct summand, then* (*trivial boundary on* $H(C; G)$)

$$H(W \otimes_\pi C; G) \cong H(W \otimes_\pi H(C; G))$$

under a natural isomorphism.

Applying these lemmas to our situation we get:

LEMMA 2.3. *If* μ: $X \to Y$ *induces* μ_*: $H(X) \cong H(Y)$, *then*

$$\mu_{\pi_*}\colon H(W^{(n)} \otimes_\pi C(X)^p; G) \cong H(W^{(n)} \otimes C(Y)^p; G).$$

LEMMA 2.4. *If $G=K$ is a field or if $G=Z$, $H(X,Z)$ has no torsion and the cycles of $C(X)$ are a direct summand, then*

$$H(W^{(n)}\otimes_\pi C(X)^p;G)\cong H(W^{(n)}\otimes_\pi H(X;G)^p)$$

under a natural isomorphism.

Definition 2.1. Now consider the case where p is an odd prime, π is the cyclic group of order p with generator α. Then we take for W the complex:

$$\cdots\to Z\pi\xrightarrow{\Gamma}Z\xrightarrow{\Delta}Z\pi\to\cdots\to Z\pi\xrightarrow{\Gamma}Z\pi\xrightarrow{\Delta}Z\pi\xrightarrow{\epsilon}Z.$$

where $Z\pi$ is the group ring of π; i.e., W_i has a single π-generator e_i, $i\geqq 0$, and

$$\begin{aligned}&\partial e_{2i+1}=\Delta e_{2i}, && \Delta=\alpha-1,\quad i\geqq 0\\ &\partial e_{2i+2}=\Gamma e_{2i+1}, && \Gamma=1+\alpha+\cdots+\alpha^{p-1},\quad i\geqq 0.\\ &\epsilon e_1=1.\end{aligned}$$

From this and Lemma 2.4 we get:

PROPOSITION 2.2. *Let π be the cyclic group on p elements, p an odd prime, W the free π-resolution of Z defined above, and $K=H(X;Z_p)$ a mod p chain complex with trivial boundary. Then if $x_1,x_2,\cdots$ is a vector basis of homogeneous elments for K, the homology classes of the following cycles form a vector basis for $H(W\otimes_\pi K^p)$:*

$$e_i\otimes_\pi x^p{}_j,\ \text{all } j,\ i>0,\ x^p{}_j=x_j\otimes\cdots\otimes x_j\ (p\ \text{times})$$
$$e_0\otimes_\pi x_{j_1}\otimes\cdots\otimes x_{j_p},$$

where we choose one representative from each class obtained by cyclic permutation of the indices.

Proof. In the proof we will represent chains in $W\otimes_\pi K^p$ by chains in $W\otimes K^p$ but consider them as identified under π.

Let $c_n=\sum_{k+l=n}e_k\otimes d_l$, $d_l\in K^p$, be any chain of $W\otimes K^p$. Then $\partial c_n=\sum\partial e_k\otimes d_l$ and hence $\partial c_n=0$ implies

(a) K even (and not zero)

$$\partial e_k\otimes d_l=\Gamma e_{k-1}\otimes d_l=e_{k-1}\otimes\Gamma d_l=0$$

for each l, and therefore

$$\Gamma d_l=0 \bmod p.$$

(b) K odd

$$\partial e_k \otimes d_l = \Delta e_{k-1} \otimes d_l = e_{k-1} \otimes \Delta d_l = 0$$

for each l, and

$$\Delta d_l = 0 \bmod p.$$

Further, all chains of the following forms are cycles:

(a) $e_k \otimes d_l$, k even $(\neq 0)$, $\Gamma d_l = 0 \bmod p$

(b) $e_k \otimes d_l$, k odd, $\Delta d_l = 0 \bmod p$

(c) $e_0 \otimes d_l$, d_l arbitrary.

We proceed to examine these cases:

(a) $\Gamma d_l = 0 \bmod p$, k even

d_l is a linear combination of terms of the form $x_{j_1} \otimes \cdots \otimes x_{j_p}$.

Divide the terms up into transitivity classes under π; i. e. $d_l = d^1{}_l + \cdots + d^t{}_l$ with $\Gamma d^r{}_l$ linearly independent from the other $\Gamma d^s{}_l$, $s \neq r$. Hence $\Gamma d^r{}_l = 0$. But $d^r{}_l = P(\alpha)(x_{j_1} \otimes \cdots \otimes x_{j_p})$ for some fixed term $x_{j_1} \otimes \cdots \otimes x_{j_p}$, where $P(\alpha) \in Z_p(\pi)$; and either $x_{j_1} \otimes \cdots \otimes x_{j_p} = x^p{}_j$ or the representation $Z_p(\pi) \to Z_p(\pi)(x_{j_1} \otimes \cdots \otimes x_{j_p})$ is faithful. In the first case, $e_k \otimes d^r{}_l$ is simply a multiple of $e_k \otimes x^p{}_j$ as demanded by the proposition. In the second case, we must have $\Gamma P(\alpha) = 0$ in $Z_p(\pi)$, and it follows (by the well-known argument for the acyclicity of W) that $P(\alpha) = \Delta Q(\alpha)$, for some $Q(\alpha) \in Z_p(\pi)$, and $e_k \otimes d^r{}_l = \partial(Q(\alpha) e_{k+1} \otimes d^r)$.

(b) $\Delta d_l = 0 \bmod p$, k odd

Writing $d_l = d^1{}_l + \cdots + d^t{}_l$ as in (a), we have

$$\Delta d^r{}_l = \Delta P(\alpha)(x_{j_1} \otimes \cdots \otimes x_{j_p}) = 0.$$

Again either $x_{j_1} \otimes \cdots \otimes x_{j_p} = x^p{}_j$ or $\Delta P(\alpha) = 0$. Now $\Delta P(\alpha) = 0$ implies $P(\alpha) = \Gamma Q(\alpha)$ and $e_k \otimes d^r{}_l = \partial(Q(\alpha) e_{k+1} \otimes d^r{}_l)$. Hence again either $e_k \otimes d^r{}_l$ is a boundary or it is a multiple of $e_k \otimes x^p{}_j$.

(c) Since

$$\partial(e_1 \otimes x_{j_1} \otimes \cdots \otimes x_{j_p}) = \Delta e_0 \otimes x_{j_1} \otimes \cdots \otimes x_{j_p} = e_0 \otimes \Delta(x_{j_1} \otimes \cdots \otimes x_{j_p}),$$

and $\Delta = \alpha - 1$; we see that two cycles of the form $e_0 \otimes x_{j_1} \otimes \cdots \otimes x_{j_p}$ are homologous if and only if they differ by a cyclic permutation of indices.

Finally we note that no cycle of the form $e_i \otimes x^p{}_j$ can be a boundary; since it only could be a boundary of something of the form $P(\alpha)e_{i+1} \otimes x^p{}_j$. But $\partial(P(\alpha)e_{i+1} \otimes x^p{}_j) = P(\alpha)\partial e_{i+1} \otimes x^p{}_j = 0$.

This proves the proposition.

Remark. If instead of $W \otimes_\pi K^p$ we consider $W^{(n)} \otimes_\pi K^p$, $n > 0$, then it is clear from the above proof, that we get the same result except that for the first class of cycles in Proposition 2.2 we restrict i so that $0 < i \leqq n$, and in addition we get all cycles of the form

$$e_n \otimes P(\alpha)(x_{j_1} \otimes \cdots \otimes x_{j_p}), \qquad P(\alpha) = 0,\ n \text{ even}$$

$$\Delta P(\alpha) = 0,\ n \text{ odd}$$

Consider now the map: $\Theta_\pi = \theta_\pi h_\pi F_\pi \colon H(W^{(n)} \otimes_\pi C(X)^p, G) \to H(X; G)$.

Definition 2.2. Let X be an $H^n{}_p$-space. Let $x \in H_j(X; Z_p)$, For $0 \leqq i \leqq n$, define $Q_i(x) = Q_i^{(p)}(x) \in H_{pj+i}(X, Z_p)$ by

$$Q_i(x) = \Theta_\pi(e_i \otimes_\pi x^p),$$

where we write again $e_i \otimes_\pi x^p$ for the homology class in $H(W^{(n)} \otimes_\pi C(X)^p; Z_p)$ represented by this cycle. $Q_i(x)$ is called the *i-th extended p-th power of x.*

Remarks.

1. It follows from (d) of Theorem 2.1, that

$$\Theta_\pi(e_0 \otimes_\pi x_{j_1} \otimes \cdots \otimes x_{j_p}) = \theta h F(e_0 \otimes x_{j_1} \otimes \cdots \otimes x_{j_p}) = x_{j_1} x_{j_2} \cdots x_{j_p},$$

the Pontrjagin product of the homology classes. In particular,

$$Q_0(x) = x \cdot x \cdots x = x^p.$$

2. It follows from (b) of Theorem 2.1, that $Q_i(x)$, $0 \leqq i < n$, is uniquely defined (i.e. independent of the choice of F). However, $Q_n(x)$ is not determined until a choice of $t\colon W^{(n)} \to J^{(n+1)}\Sigma_p$ has been made. See Lemma 2.5 below.

3. The classes $e_n \otimes P(\alpha)(x_{j_1} \otimes \cdots \otimes x_{j_p})$ can be thought of as defining 'non-stable' operations in the sense that they disappear under an $H^n{}_p$-map of X into an H^{n+1}-space.

4. Using different coefficients in $H(W^{(n)} \otimes_\pi C(X)^p; G)$, other stable

and non-stable operations may be defined. In the non-stable case, Browder [3] has studied non-stable operations for $p = 2$, corresponding to the cycles

$$e_n \otimes \Delta(x \otimes y),\ n \text{ even}$$
$$e_n \otimes \Gamma(x \otimes y),\ n \text{ odd},$$

where x and y are cycles in X, arbitrary coefficients. In particular,

$$\text{let } \psi_n(x, y) = \psi^{(p)}_n(x, y) = \Theta_\pi(e_n \otimes_\pi \Delta(x \otimes y)),\ n \text{ even}$$
$$= \Theta_\pi(e_n \otimes_\pi \Gamma(x \otimes y)),\ n \text{ odd},$$

$x \in H_i(X; Z_p)$, $y \in H_j(X, Z_p)$. Then $\psi_n(x, y) \in H_{i+j+n}(X; Z_p)$.

In this paper, we will consider only the extended p-th powers, and the operations ψ_n defined above.

Note that $\psi_0(x, y) = \Theta_\pi(e_0 \otimes_\pi \Delta(x \otimes y)) = xy - (-1)^{ij}yx$, for an $H^0{}_p$-space X.

5. If X is connected and is a special $H^n{}_p$-space, and $1 \in H_0(X; Z_p)$ is the unit element, then $Q_i(1) = 0$, $i > 0$. In fact the extra condition implies that the inclusion of $(e) \to X$ is an $H^n{}_p$-map, (e) the space consisting of the unit element of X only. Hence $Q_i(1)$ is in the image of $H_i((e); Z_p) = 0$, $i > 0$.

LEMMA 2.5. *Let W be as in Definition 2.1 and consider $J^\infty\pi$ as a simplicial complex then there exists an equivariant chain map $S\colon W \to J^\infty\pi$, $\pi = Z_p$, such that $S(W^{(n)}) \subset J^{n+1}\pi$.*

Proof. Given a simplex $(\sigma_0(i_0), \cdots, \sigma_r(i_r))$ in $J^\infty Z_p$, we will let $(\sigma_0(i_0), \cdots, \sigma_r(i_r)) \circ \sigma_{r+1}(i_{r+1}) = (\sigma_0(i_0), \cdots, \sigma_r(i_r), \sigma_{r+1}(i_{r+1}))$,

and extend this operation linearly to chains.

Now set $S(e_0) = (\epsilon(0))$, ϵ the unit of Z_p

$$S(e_1) = (\epsilon(0), \epsilon(1)) - (\alpha(0), \epsilon(1)) = -(\Delta S e_0) \circ \epsilon(1)$$

and in general

$$S(e_{2i+1}) = (-\Delta S e_{2i}) \circ \epsilon(2i + 1)$$
$$S(e_{2i+2}) = (\Gamma S e_{2i+1}) \circ \epsilon(2i + 2)$$

Then
$$\partial S(e_{2i+1}) = (-\partial \Delta S e_{2i}) \circ \epsilon(2i + 1) + \Delta(e_{2i})$$
$$= (-\Delta \partial S e_{2i}) \circ \epsilon(2i + 1) + \Delta(e_{2i})$$
$$= (-\Delta \Gamma S e_{2i}) \circ \epsilon(2i + 1) + \Delta(e_{2i})$$
$$= \Delta S(e_{2i}).$$

Similarly $\partial S(e_{2i+2}) = \Gamma S(e_{2i+1})$.

COROLLARY. *Let* $t\colon W \to C(J^\infty Z_p)$ *be the equivariant chain map defined by sending* W *into the simplicial complex of* $J^\infty Z_p$ *as described in the proof above, then the simplicial complex into the simplicial singular complex, and finally the simplicial singular complex into the cubical singular complex by the canonical maps described in* ([5]). *Then* $t\colon W^{(n)} \to C(J^{n+1}Z_p)$ *is an equivariant chain map, and if* F *of Theorem* 2.1, *Corollary, satisfies* 1) *and* 2) *for this* t,

$$F\colon W^{(n)} \otimes C(X)^p \to C(J^{n+1}\pi) \otimes C(X^p),$$

Any two maps F, F^1 *with equivariant chain map* t, *are equivariantly chain homotopic.*

If we define the operation $Q^{(p)}_n$ using this t and such an F, it will be uniquely defined in an H^n_p-space. Likewise for the Browder operations $\psi^{(p)}_n$.

THEOREM 2.2. *Let* X *be an* H^n_p*-space. Then*

a) $Q^{(p)}_i$ *is a homomorphism for* $i \leqq n-1$.

b) $Q^{(p)}_0$ *is the Pontrjagin p-th power.*

c) *If* ∂_p *is the homology Bockstein operator induced by the sequence* $0 \to Z_p \to Z_{p^2} \to Z_p \to 0$, *then* $Q^{(p)}_{2i-1} = \partial_p Q^{(p)}_{2i}$, $2i \leqq n-1$.

d) *For* p *an odd prime,* $x \in H_r(X;Z_p)$, $Q^{(p)}_{2i}(x) = 0$ *unless the change in dimension,* $2i + pr - r$ *is an even multiple of* $p-1$.

e) *The operations are natural under* H^n_p*-maps.*

Proof.

(a) For $i < n$, we show that for $x, y \in H_r(X;Z_p)$,

$$T = (e_i \otimes_\pi (x+y)^p) - (e_i \otimes_\pi x^p) - (e_i \otimes_\pi y^p),$$

is a boundary in $W^{(n)} \otimes_\pi H(X;Zp)^p$. This will give (a) above.

Since $(x+y)^p - x^p - y^p$ is invariant under π, the terms in its expansion are divided into transivity classes. But after cancelling out x^p and y^p all remaining terms in $(x+y)^p$ involve both x's and y's, and cannot be left fixed by any cyclic permutation. Consequently, $T = \Gamma S$, where S is obtained by choosing one representative from each transitivity class in T. Since $\Delta\Gamma = 0$ and $\Gamma\Gamma = p\Gamma = 0 \bmod p$, it follows as in proof of Proposition 2.2, that T is a boundary. Explicitly, since $\Gamma = \Delta^{p-1} \bmod p$,

$$\partial(e_{i+1} \otimes S) = T \quad (i \text{ odd}), \qquad \partial(e_{i+1} \otimes \Delta^{p-2}S) = T \quad (i \text{ even})$$

(b) This follows from the definition (see Remark 1 on Definition 2.3).

(c) In $W^{(n)} \otimes_\pi C(X)^p$, consider $e_{2i} \otimes_\pi x^p$ where x is a mod p cycle; i. e. $\partial x = py$. Then

$$\begin{aligned}\partial(e_{2i} \otimes_\pi . x^p) &= e_{2i-1} \otimes_\pi x^p + p \sum_k (-1)^{\dim x(k-1)} e_{2i} \otimes_\pi x^{k-1} y^{p-k} \\ &= pe_{2i-1} \otimes_\pi x^p + pe_{2i} \otimes_\pi \Gamma(yx^{p-1}) \\ &= p\{e_{2i} \otimes_\pi x^p + \partial(e_{2i+1} \otimes_\pi x^{p-2}(yx^{p-1})\} \bmod p^2.\end{aligned}$$

The result follows.

(d) From the definitions we have the commutative diagram:

$$\begin{tikzcd} W^{(n)} \otimes C(X)^p \arrow[r, "t \otimes i^p"] & J^{n+1}\pi \otimes C(X)^p \arrow[r, "\theta \circ F"] \arrow[d, "j \otimes i^p"] & C(X) \\ & J^{n+1}\Sigma_p \otimes C(X)^p \arrow[ur, "\theta \circ \bar{F}"'] & \end{tikzcd}$$

where i^p: $C(X)^p \to C(X)^p$ is the identity, j: $J^{n+1}\pi \to J^{p+1}\Sigma_p$ is the inclusion induced by the inclusion of π in Σ_p, F, $\bar{F}$ are the maps of Proposition 2.1, Corollary for π, Σ_p, respectively, and θ is the defining map for the $H^n{}_p$ structure of X. All maps are π equivariant, but $\theta \circ \bar{F}$ is Σ_p equivariant as well.

From the above we get

$$\begin{tikzcd} (W^{(n)} \otimes Z_p) \otimes_\pi C(X)^p \otimes Z_p \arrow[r, "(\theta \circ F)_\pi"] \arrow[d, "j \circ t \otimes_\pi i^p"] & C(X) \otimes Z_p \otimes_{\Sigma_p} Z_p \\ (J^{n+1}\Sigma_p \otimes Z_p) \otimes_{\Sigma_p} (C(X)^p \otimes Z_p) \arrow[ur, "(\theta \circ \bar{F})_{\Sigma_p}"'] & \end{tikzcd}$$

commutes in the following cases:

i) Σ_p acts trivially on both coefficient groups Z_p

ii) Each $\sigma \in \Sigma_p$ multiplies both coefficients by the sign of σ.

In both cases, π acts trivially on the coefficients since cyclic permutation, p odd prime, are even. Also $Z_p \otimes_{\Sigma_p} Z_p \cong Z_p$ with trivial action. Hence in either case we have:

$$\begin{tikzcd} H(W^{(n)} \otimes_\pi C(X)^p ; Z_p) \arrow[r, "\theta_*"] \arrow[dr, "(j \circ t \otimes_\pi i^p)"] & H(X ; Z_p) \\ & H(J^{p+1}\Sigma_p \otimes Z_p) \otimes_{\Sigma_p} (C(X)^p \otimes Z_p)) \arrow[u, "(\theta \circ \bar{F})_{\Sigma_p}"'] \end{tikzcd}$$

commutes.

Let x is a cycle in $C(X)$, thus $j\circ t\otimes_{\pi} i^{p}(e_{2i}\otimes_{\pi} x^{p}) = j\circ t(e_{2i})\otimes_{\Sigma_p} x^{p}$. Let $\bar{e}_{2i} = j\circ t(e_{2i})$. Thus since e_{2i} represents a mod p cycle in $W^{(n)}/\pi$, $\bar{e}_{2i}$ represents a mod p cycle in $J^{n+1}\Sigma p/\Sigma p$. Consider

i) x even dimensional

If $\bar{e}_{2i} = \partial c \bmod p$ in $J^{n+1}\Sigma p/\Sigma p$, then

$$\bar{e}_{2i}\otimes_{\Sigma_p} x^{p} = \partial(c\otimes_{\Sigma_p} x^{p}) \bmod p,$$

Since $\partial c\otimes_{\Sigma_p} x^{p}$ depends only on the class of ∂c under Σ_p as the action of Σp on x^{p} is trivial. It will then follow from the above diagram, case i, that $\Theta(e_{2i}\otimes_{\pi} x^{p})$ is homologous to zero and hence $Q^{(p)}{}_{2i}(x) = 0$.

ii) odd dimensional

If $\bar{e}_{2i} = \partial c$ in $J^{n+1}\Sigma p\otimes_{\Sigma_p} Z_p$ acting in Z_p by sign of permutation; then $\bar{e}_{2i}\otimes_{\Sigma_p} x^{p} = \partial(c\otimes x^{p})$ in $(J^{n+1}\Sigma_p\otimes Z_p)\otimes_{\Sigma_p}(C(X)^{p}\otimes Z_p)$, as the action of Σp on x^{p} is trivial in $C(X)^{p}\otimes Zp$; Σ_p acting on Z_p by the sign of the permutation. Hence again, $Q^{(p)}{}_{2i}(x) = 0$. Now observe that $H_i(\pi) = H_i(W^{(r)}/\pi)$, $i < n$; and $H_i(\Sigma_p) = H_i(J^{n+1}\Sigma p/\Sigma p)$, $i < n$; and further that

$$H_i(W^{(n)}/\pi) \xrightarrow{(J\circ t)_{\Sigma_p}} H_i(J^{n+1}\Sigma_p/\Sigma_p)$$

corresponds to the inclusion map of π in Σp. Hence our result will follow from:

LEMMA. *Let* $j_{2i}\colon H_{2i}(\pi; Zp)\to H_{2i}(\Sigma_p; ZZ_p)$ *be induced by the inclusion map; then*

i) *If* Σp *acts trivially on* Zp, $j_{2i} = 0$ *if* $2i$ *is not an even multiple of* $p-1$.

ii) *If* Σ_p *acts on* Z_p *by the sign of the permutation,* $j_{2i} = 0$ *if* $2i$ *is not an odd multiple of* $p-1$. (For proof see [15]).

(e) This follows from the definitions.

Remark. If x is an integral cycle, then $e_{2i-1}\otimes x^{p}$ is an integral cycle, and we may consider Q_{2i-1} as an integral operation. However, letting ∂_{*} be the Bockstein for the sequence $0\to Z\to Z\to Zp\to 0$, we get again as in (c) that:

$$Q_{2i-1} = \partial_{*}Q_{2i}.$$

We now derive a Cartan formula for operations on products:

For two H^n_p-spaces X and Y, $X \times Y$ has an obvious H^n_p-structure given by:

$$\theta\colon J^{n+1}\Sigma_p \times (X \times Y)^p \to X \times Y,$$
$$\theta(w;(x_1,y_1),\cdots,(x_p,y_p)) = (\theta_x(w;x_1,\cdots,x_p),\theta_y(w;y_1,\cdots,y_p)),$$

where θ_x and θ_y are the defining maps for X and Y respectively. The properties for θ follow immediately from those for θ_x, θ_y. The same formula holds if we restrict to $J^{n+1}\pi$, $\pi \subset \Sigma_p$. We may write θ as the composition of maps:

$$J^{n+1}\pi \times (X \times Y)^p \xrightarrow{\Delta \times i} J^{n+1}\pi \times J^{n+1}\pi \times (X \times Y)^p$$
$$\xrightarrow{\tau} J^{n+1}\pi \times X^p \times J^{n+1}\pi \times Y^p \xrightarrow{\theta_x \times \theta_y} X \times Y,$$

where Δ is the diagonal map on $J^{n+1}\pi$, τ the permutation of factors.

In the following diagram we write simply J for $J^{n+1}\pi$; the bottom line is the chain map induced by θ.

$$\begin{tikzcd}
C(J)\otimes C((X\times Y)^p) \arrow[r, "\Delta\otimes i"] \arrow[d] & C(J\times J)\otimes C((X\times Y)^p) \arrow[r, "G_1\times G_2"] \arrow[d] & C(J)\otimes C(J)\otimes C(X^p)\otimes C(Y^p) \cdot \\
C(J\times(X\times Y)^p) \arrow[r, "\Delta\times i"] & C(J\times J\times(X\times Y)^p) \arrow[r, dash] & {}
\end{tikzcd}$$

$$\begin{tikzcd}
C(J)\otimes C(X^p)\otimes C(J)\otimes C(Y^p) \arrow[r, "\theta_x\otimes\theta_y"] \arrow[d] & C(X)\otimes C(Y) \arrow[d] \\
C(J\times X^p\times J\times Y^p) \arrow[r, "\theta_x\times\theta_y"] & C(X\times Y).
\end{tikzcd}$$

Here G_1, G_2 are the natural maps from direct product to tensor product given by the Eilenberg-Zilber theorem, and the vertical maps are the natural chain maps from tensor product to direct product. The fact that the above squares all commute up to π-equivariant chain homotopies, follows from the fact that they all commute up to natural chain homotopies and π operates on each factor separately. (I. e. π induces a map in the category).

Consider next the diagram

$$\begin{tikzcd}
C(J)\otimes(C(X)\otimes C(Y))^p \arrow[r, "G_1\Delta\otimes i"] \arrow[d, "i\otimes f^p"] & C(J)\otimes C(J)\otimes(C(X)\otimes C(Y))^p \arrow[r, "i\otimes\sigma"] & C(J)\otimes C(J)\otimes C(X)^p\otimes C(Y)^p \arrow[d, "\tau^{-1}F_x\otimes F_y\tau"] \\
C(J)\otimes C(X\times Y)^p \arrow[r, "F_{x\times y}"] & C(J)\otimes C((X\otimes Y)^p) \arrow[r, "(G_1\otimes G_2)\circ(\Delta\otimes i)"] & C(J)\otimes C(J)\otimes C(X^p)\otimes C(Y^p)
\end{tikzcd}$$

$$\begin{tikzcd}
{} \arrow[r, "\tau"] & C(J)\otimes C(X)^p\otimes C(J)\otimes C(Y)^p \arrow[d, "F_x\otimes F_y"] \\
{} \arrow[r, "\tau"] & C(J)\otimes C(X^p)\otimes C(J)\otimes C(Y^p).
\end{tikzcd}$$

The maps F_x and F_y are chosen to satisfy Theorem 2.1 for the standard maps $C(X)^p\to C(X^p)$ and $C(Y)^p\to C(Y^p)$ respectively. The second square then commutes by definition. Also $F_{x\times y}$ is chosen to satisfy Theorem 2.1 for the standard map $C(X\times Y)^p\to C((X\times Y)^p)$. Further, $f\colon C(X)\otimes C(Y)\to C(X\times Y)$ is the standard map. Finally, σ is the permutation of factors. The first square commutes up to a π-equivariant homotopy by Theorem 2.1, because in going around either way in the square we get a map satisfying Theorem 2.1 for the map: $(x_1\otimes y\ )\cdots(x_p\otimes y_p)\to(x_1*\cdots*y_p)\otimes(y_1*\cdots*y_p)$ where x_i and y_i, $i=1,\cdots,p$, are zero dimensional.

Combining the two diagrams we have:

$$\begin{tikzcd}
C(J)\otimes(C(X)\otimes C(Y))^p \arrow[r, "G_1\Delta\otimes i"] \arrow[d, "i\otimes f^p"] & C(J)\otimes C(J)\otimes(C(X)\otimes C(Y))^p \arrow[r, "\tau(i\otimes\sigma)"] & C(J)\otimes C(X)^p\otimes C(J)\otimes C(Y)^p \arrow[d, "\Theta_x\otimes\Theta_y"] \\
C(J)\otimes C(X\times Y)^p \arrow[r, "\Theta"] & C(X\times Y) & C(X)\otimes C(Y) \arrow[l, "f"']
\end{tikzcd}$$

commutes up to a π-equivariant homotopy.

Now let W by the π-free acyclic complex defined previously (Definition 2.1). If we restrict to $W^{(n+1)}$, any two π-homotopic, since $C(J^{n+1}\pi)\otimes C(J^{n+1}\pi)$ is acyclic in dimensions less than n.

Consequently, we have

LEMMA 2.6.

$$\begin{tikzcd}
W^{(n-1)}\otimes(C(X)\otimes C(Y))^p \arrow[r, "d\otimes i"] \arrow[d, phantom, "\Theta\circ i\otimes f^p"] & W^{(n-1)}\otimes W^{(n-1)}\otimes(C(X)C(Y))^p \arrow[r, "\tau(i\otimes\sigma)"] & W^{(n-1)}\otimes C(X)^p\otimes W^{(n-1)}\otimes C(Y)^p \arrow[d, "\Theta_x\otimes\Theta_y"] \\
C(X\times Y) & & C(X)\otimes C(Y) \arrow[ll, "f"']
\end{tikzcd}$$

commutes up to a π-equivariant homotopy.

We are now in a position to prove the following:

THEOREM 2.3. (a) *For p an odd prime, X and Y $H^n{}_p$-spaces, $x_r \in H_r(X;Z_p)$, $y_s \in H_s(Y;Z_p)$, $2i < n$, the following holds in $X \times Y$:*

$$Q^{(p)}{}_{2i}(x_r \otimes y_s) = (-1)^{rs(p-1/2)} \sum_{j=0}^{i} Q^{(p)}{}_{2j}(x_r) \otimes Q^{(p)}{}_{2i-2j}(y_s)$$

(b) *For $p = 2$, $i < n$*

$$Q^{(2)}{}_i(x_r \otimes y_s) = \sum_{j=0}^{i} Q^{(2)}{}_j(x_r) \otimes Q^{(2)}{}_{i-j}(y_s).$$

COROLLARY 1. *If X is the n-th loop space of an H-space, then the following holds where multiplication indicates Pontrjagin product:*

$$Q^{(p)}{}_{2i}(x_r \cdot y_s) = (-1)^{rs(p-1/2)} \sum_{j=0}^{i} Q^{(p)}{}_{2j}(x_r) \cdot Q^{(p)}{}_{2i-2j}(y_s), \quad 2i < n$$

$$Q^{(2)}{}_i(x_r \cdot y_s) = \sum_{j=0}^{i} Q^{(2)}{}_j(x_r) \cdot Q^{(2)}{}_{i-j}(y_s), \quad i < n$$

where $x_r \in H_r(X;Z_p)$, $y_s \in H_s(X;Z_p)$.

Proof. Let π be the cyclic group on p elements with generator α, W the π-free acyclic complex of Definition 2.1; let as in [13], $d\colon W \to W \otimes W$ be defined by,

$$de_{2i} = \sum_{j=0}^{i} e_{2j} \otimes e_{2i-2j} + \sum_{j=0}^{i-1} \Lambda e_{2j+1} \otimes e_{2i-2j-1}$$

and

$$de_{2i+1} = \sum_{j=0}^{i} (e_{2j} \otimes e_{2i-2j+1} + e_{2j+1} \otimes \alpha e_{2i-2j}),$$

where $\Lambda = \sum \alpha^k \times \alpha^l$, $0 \leqq k < l \leqq p-1$. Then d is an equivariant chain map.

Using the commutative diagram of Lemma 2.6 and mod p coefficients, we have:

$$\Theta_\pi(e_{2i} \otimes_\pi (x_r \otimes y_s)^p) = f \circ (\Theta_x)_\pi \otimes (\Theta_y)_\pi \circ \tau \circ (i \otimes \sigma)(de_{2i} \otimes_\pi (x_r \otimes y_s)^r)$$

$$= f \circ (\Theta_x)_\pi \otimes (\Theta_y)_\pi \{ (-1)^{p(p-1/2)rs} \sum_{j=0}^{i} (e_{2j} \otimes_\pi x^p{}_r) \otimes (e_{2i-2j} \otimes_\pi y^p{}_s)$$

$$+ \sum_{j=0}^{i-1} (-1)^{(2i-2j-1)rp} \sum_{k<l} (\alpha^k e_{2j+1} \otimes_\pi x^p{}_r) \otimes (\alpha^l e_{2i-2j-1} \otimes_\pi y^p{}_s).$$

But $x^p{}_r$ and $y^p{}_s$ are invariant under π and hence for p odd prime:

$$\sum_{k<l} (\alpha^k e_{2j+1} \otimes_\pi x^p{}_r) x (\alpha^l e_{2i-2j-1} \otimes_\pi y^p{}_s)$$
$$= p(p-1/2)(e_{2j+1} \otimes_\pi x^p{}_r) \otimes (e_{2i-2j-1} \otimes_\pi y^p{}_s) = 0$$

Hence

$$\Theta_\pi(e_{2i}\otimes_\pi(x_r\otimes y_s)^p)$$
$$= (-1)^{(p-1/2)r\cdot s} f\sum_{j=0}^{i}(\Theta_x)_\pi(e_{2j}\otimes_\pi x^p{}_r)\otimes(\Theta_y)_\pi(e_{2i-2j}\otimes_\pi y^p{}_s)$$

or

$$Q^{(p)}{}_{2i}(x_r\otimes y_s) = (-1)^{(p-1/2)rs}\sum_{j=0}^{i}Q^{(p)}{}_{2j}(x_r)\otimes Q^{(p)}{}_{2i-2j}(y_s).$$

Similarly, for $p=2$, using the above result and the corresponding one for e_{2i+1}, we get for i odd or even:

$$\Theta_\pi(e_i\otimes_\pi(x_r\otimes y_s)^p) = f\sum_{j=0}^{i}(\Theta_x)_\pi(e_j\otimes_\pi x^p{}_r)\otimes(\Theta_y)_\pi(e_{i-j}\otimes_\pi y^p{}_s)$$

or

$$Q^{(2)}{}_i(x_r\otimes y_s) = \sum_{j=0}^{i}Q^{(2)}{}_j(x_r)\otimes Q^{(2)}{}_{i-j}(y_s).$$

This proves the theorem.

Proof of Corollary 1. Now assume $X=\Omega^n Z$, where Z is an H-space, $h\colon Z\times Z\to Z$ the multiplication. This induces $\Omega^n h\colon \Omega^n(Z\times Z)\to\Omega^n Z$. On the other hand, the H-structure in X is defined by considering $X=\Omega(\Omega^{n-1}Z)$, and adding loops to define the multiplication $\mu\colon X\times X\to X$.

We wish to apply the theorem to the case where $Y=X$. Since X is an $H^n{}_p$-space, $X\times X$ has an $H^n{}_p$-structure as defined previously. But, $\Omega^n(Z\times Z)$ has an H^{n-1}-structure, and it is easy to check that the canonical map $i\colon \Omega^n(Z\times Z)\to\Omega^n Z\times\Omega^n Z$ is an $H^{n-1}{}_p$-map. Further, it is well known and trivial to prove that

$$\begin{tikzcd}
\Omega^n(Z\times Z) \arrow[rr, "i"] \arrow[dr, "\Omega^n h"'] & & \Omega^n Z\times\Omega^n Z \arrow[dl, "\mu"] \\
& \Omega^n Z &
\end{tikzcd}$$

commutes up to homotopy. Since $i_*\colon H_*(\Omega^n(Z\times Z);Z_p)\cong H_*(\Omega^n Z\times\Omega^n Z)$ and $\Omega^n h$ is an $H^{n-1}{}_p$-map, it follows that μ_* commutes with the homology operations:

$$\mu_* Q^p{}_{2i} = Q^{(p)}{}_{2i}\mu_* \quad 2i<n; \qquad \mu_* Q^{(2)}{}_i = Q^{(2)}{}_i\mu_*, \quad i<n$$

the corollary now follows immediately by applying μ_* to both sides of the equations of the theorem.

COROLLARY 2. *If X is an $H^n{}_p$-space, and $x_r \in H_r(X;Z_p)$ is a primitive homology class; then $Q_i(x_r)$ is primitive, $i < n$, p any prime.*

Proof. The diagonal map $X \rightarrow X \times X$ is an $H^n{}_p$-map, and hence commutes with the operations. The result now follows from the theorem.

III. Further properties of homology operations. As seen in Theorem 2.2(d) many of the homology operations $Q^{(p)}{}_i$ are trivial. It is thus convenient to define

$$S^j{}_{(p)}(x) = Q^{(p)}{}_{(2j-n)(p-1)}(x), \quad n = \dim x.$$

We shall find it convenient to introduce also a numerical coefficient.

THEOREM 3.1. *Let $\sigma_*\colon H_{n-1}(\Omega X;Z_p) \rightarrow H_n(X;Z_p)$ be the usual homology suspension and suppose for $y \in H_{n-1}(\Omega X;Z_p)$ both $S^j{}_{(p)}(y)$ and $S^j{}_{(p)}(\sigma_* y)$ are defined (this requires X to be an $H^m{}_p$-space, where*

$$m+1 \geqq (2j-n+1)(p-1)),$$

then

$$\sigma_* S^j{}_{(p)}(y) = (-1)^{(p-1/2)} \cdot \gamma(n) \cdot S^j{}_{(p)}(\sigma_* y),$$

where $\gamma(n) \equiv (-1)^{(n-1)(p-1)/2} \cdot ((p-1)/2\,!) \bmod p$.

Proof. Let PX denote the space of paths over X and $\pi\colon PX \rightarrow X$ denote the end point projection. Suppose given classes

$$x \in H_n(X;Z_p),\ y \in H_{n-1}(\Omega X;Z_p) \text{ and } z \in H_n(PX,\Omega X;Z_p)$$

such that

$$\pi_*(z) = x \in H_n(X,e;Z_p) = H_n(X;Z_p) \text{ and } \partial_*(z) = y.$$

We must show there is a chain $c \in C_{pn+i}(PX)$, where $i = (2j-n)(p-1)$, such that

$$\begin{aligned}&\{c\} \in H_{pn+i}(PX,\Omega X;Z_p),\\ &\pi_*\{c\} = (-1)^{(p-1)/2} \cdot \gamma(n) \cdot S^j{}_{(p)}(x), \text{ and}\\ &\partial_*\{c\} = S^j{}_{(p)}(y).\end{aligned}$$

Let (z,y), $y = \partial z$, denote the subchain complex of $C(PX)$, the normalized cubical singular complex of PX, generated by a chain representing z (which we also denote by z) and by ∂z, which represents y.

As X is an $H^m{}_p$-space, we have the equivariant maps

$$\begin{aligned}&\theta^m{}_p\colon J^{m+2}\Sigma_p \times X^p \rightarrow X \text{ and}\\ &\theta^{m+1}{}_p\colon J^{m+2}\Sigma_p \times (\Omega X)^p \rightarrow \Omega X,\end{aligned}$$

the latter being defined by formula (1.4).

Let $f\colon J^{m+1}\Sigma_p \to J^{m+2}\Sigma_p$ be defined by

$$f(t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m) = (t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m \oplus 0)$$

and let

$$\widetilde{\theta^m}_p\colon J^{m+2}\Sigma_p \times (\Omega X)^p \to \Omega X$$

be the composition $\theta^{m+1}_p \circ (f \times (\mathrm{id.})^p)$.

Define

$$\bar{\theta}^m_p\colon J^{m+1}\Sigma_p \times (PX)^p \to PX$$

by

$$\bar{\theta}^m_p(t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m; w_1, \cdots, w_p)(t) = \theta^m_p(t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m; w_1(t), \cdots, w_p(t))$$

and define

$$\hat{\theta}^m_p\colon (J^{m+1}\Sigma_p \circ \epsilon) \times (\Omega X)^{p-1} \times (PX) \to PX$$

by

$$\hat{\theta}^m_p((1-s)t_0\sigma_0 \oplus \cdots \oplus (1-s)t_m\sigma_m \oplus s\epsilon; l_1, \cdots, l_{p-1}, w)(t) = \theta^m_p(t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m; l^s(t), \cdots, l^s_{p-1}(t), w^s(t)),$$

where, as before,

$$l^s_i(t) = l_i(t - s(r_1 + \cdots + r_{i-1})) \text{ and } w^s(t) = w(t - s(r_1 + \cdots + r_{p-1})).$$

It is clear that

$$\pi \circ \hat{\theta}^m_p(t_0\sigma_0 \oplus \cdots \oplus t_m\sigma_m \oplus t_{m+1}\epsilon; l_1, \cdots, l_{p-1}, w) = \pi(w).$$

As each two of the maps $\tilde{\theta}^m_p$, $\bar{\theta}^m_p$ and $\hat{\theta}^m_p$ agree on the common part of their domains of definition, we have the commutative diagram

$$\begin{tikzcd}
J^{m+2}\Sigma_p \times (\Omega X)^p \arrow[r] \arrow[d] & \Omega X \arrow[d] \\
{[J^{m+2}\Sigma_p \times (\Omega X)^p] \cup [(J^{m+1}\Sigma_p)\circ\epsilon \times (\Omega X)^{p-1} \times PX] \cup [J^{m+1}\Sigma_p \times (PX)^p]} \arrow[r] & PX \arrow[dd] \\
J^{m+1}\Sigma_p \times (PX)^p \arrow[u] \arrow[d] & \\
J^{m+1}\Sigma_p \times X^p \arrow[r] & X
\end{tikzcd}$$

Let $W^{(k)}$ be the k-skeleton of the standard acyclic complex for Z_p, and let $(W^{(m+1)}, e_{m+2})$ denote the subcomplex of $W^{(m+2)}$ having only the generator e_{m+2} in dimension $m+2$ (not $T^i e_{m+2}, 1 < i < p$).

Then by the Corollary of Theorem 2.1 the diagram

$$\begin{tikzcd}
W^{(m+2)}\otimes C(\Omega X)^p \arrow[r] \arrow[d] & C(\Omega X) \arrow[d] \\
{[W^{(m+2)}\otimes C(\Omega X)^p] \oplus [(W^{(m+1)}, e_{m+2})\otimes C(\Omega X)^{p-1}\otimes C(PX)] \oplus [W^{(m+1)}\otimes C(PX)^p]} \arrow[r] & C(PX) \arrow[dd] \\
W^{(m+1)}\otimes C(PX)^p \arrow[u] \arrow[d] & \\
W^{(m+1)}\otimes C(X)^p \arrow[r] & C(X)
\end{tikzcd}$$

commutes.

The chain complex

$$\{W^{(m+2)}\otimes y^p\} \oplus \{(W^{(m+1)}, e_{m+2})\otimes y^{p-1}\otimes z\} \oplus \{W^{(m+1)}\otimes (z,y)^p\}$$

is a subcomplex of the chain complex

$$[W^{(m+2)}\otimes C(\Omega X)^p] \oplus [(W^{(m+1)}, e_{m+2})\otimes C(\Omega X)^{p-1}\otimes C(PX)] \oplus [W^{(m+1)}\otimes C(PX)^p]$$

and the image of any term of this subcomplex involving at least one factor y projects under π into a degenerate chain in $C(X)$ since $y \in C(\Omega X)$.

We shall define a chain c' in this subcomplex such that modulo residual chains and with coefficients Z_p

$$c' = (-1)^{p-1/2}\cdot\gamma(n)\cdot e_t\otimes z^p + \text{terms involving at least one } y,$$

and

$$\partial c' = e_{t+p-1}\otimes y^p.$$

Then if c is the image of this chain in $C(PX)$, it follows from the above that c has the required properties.

To define c' it is convenient when p is an odd prime to replace $(z,y)^p$ by a smaller chain complex, and to this end we use a lemma of R. G. Swan [17].

LEMMA. *Let K be the chain complex*

$$0 \to Z_p \xrightarrow{\Gamma} KZ_p \xrightarrow{\Delta} \cdots \xrightarrow{\Gamma} KZ_p \xrightarrow{\Delta} KZ_p \xrightarrow{n} Z_p \to 0,$$

where KZ_p denotes the Z_p group ring of Z_p, $p \neq 2$, n is augmentation, $\Delta = \alpha - 1$, $\Gamma = \alpha^{p-1} + \cdots + \alpha + 1$, and dimensions range from $(n-1)p$ to np. (α acts as left multiplication in KZ_p and trivially on the two end groups Z_p; this action commutes with boundary.) Then there is an equivariant equivalence $\phi\colon K \to (z,y)^p$ such that

$$\phi(\epsilon) = y^p,\ \epsilon \in Z_p = K_{(n-1)p}$$

$$\phi(\epsilon) = y^{p-1}\otimes z,\ \epsilon \in KZ_p = K_{(n-1)p+1},\ \textit{and}$$

$$\phi(\epsilon) = \gamma(n)\cdot z^p,\ \epsilon \in Z_p = K_{np},\ \textit{where}$$

$\gamma(n) \equiv (-1)^{(n-1)(p-1)/2}\cdot((p-1)/2)!)\operatorname{mod} p$.

We shall define a chain c'' in $W^{(n)} \otimes K$ such that its image in $W^{(n)} \otimes (z, p)^p$ under the map $1 \otimes \phi$ is the desired chain c'.

Since $i = (2j - n)(p - 1)$ and p is odd, i is even. We let

$$\begin{aligned} c'' = {} & e_{i+p-1} \otimes \epsilon_{(n-1)p+1} + e_{i+p-2} \otimes \Delta^{p-2} \epsilon_{(n-1)p+2} \\ & - e_{i+p-3} \otimes \epsilon_{(n-1)p+3} - e_{i+p-4} \otimes \Delta^{p-2} \epsilon_{(n-1)p+4} + \cdots \\ & + (-1)^{p-1/2} e_i \otimes \epsilon_{np}. \end{aligned}$$

Then

$$\begin{aligned} \partial c'' = e_{i+p-1} \otimes \epsilon_{(n-1)p} + {} & \Gamma e_{i+p-2} \otimes {}_{(n-1)p+1} - e_{i+p-2} \otimes \Delta^{p-1} \epsilon_{(n-1)p+1} \\ & + \Delta e_{i+p-3} \otimes \Delta^{p-2} \epsilon_{(n-1)p+2} - e_{i+p-3} \otimes \Gamma \epsilon_{(n-1)p+2} \\ & - \Gamma e_{i+p-4} \otimes \epsilon_{(n-1)+3} + \cdots \\ & + (-1)^{p-1/2} \Gamma e_{i-1} \otimes \epsilon_{np} \end{aligned}$$

$\equiv e_{i+p-1} \otimes \epsilon_{(n-1)p}$, modulo residual chains and with coefficients modulo p. (Note that in $W \otimes K$, $\alpha a \otimes b \equiv a \otimes \alpha b$, mod residual chains; hence $\Delta a \otimes b = a \otimes \Delta b$ and $\Gamma a \otimes b = a \otimes \Gamma b$. Also, $\Gamma \equiv \Delta^{p-1} \bmod p$ and $\Gamma \epsilon_{np} = 0 \bmod p$.) It is clear that c'' has the property described above.

For the case $p = 2$, it is sufficient simply to let

$$c' = e_i \otimes z^2 + e_{i+1} \otimes z \otimes y.$$

Then

$$\partial c' \equiv e_{i+1} \otimes y^2 \text{ mod } 2 \text{ and mod residual chains.}$$

As is the case with the Steenrod operations, it is also convenient here to multiply the homology operations $S^j{}_{(p)}$ by suitably selected numerical coefficients so that they will commute precisely with the homology suspension homomorphism. That is, we wish to determine numbers $f(j, n)$ such that if $Q^j{}_{(p)}(w) = f(j, n) \cdot S^j{}_{(p)}(w)$, $\dim w = n$, then

(1) $Q^k{}_{(p)}(w) = w^p$, if $n = 2k$, and

(2) $\sigma_* Q^j{}_{(p)}(w) = Q^j{}_{(p)}(\sigma_* w)$.

The conditions (1) and (2) may be translated into conditions on $f(j, n)$ as follows:

(1) $Q^k(w_{(2k)}) = w^p = S_0(w_{(2k)}) = S^k(w_{(2k)})$.

Thus, $f(k, 2k) = 1$.

(2) $$\begin{aligned} & \sigma_* Q^j(y) = \sigma_* f(j, n-1) S^j(y) = f(j, n-1) \cdot (-1)^{p-1/2} \cdot \gamma(n) \cdot S^j(\sigma_* y), \\ & \qquad \text{and } \sigma_* Q^j(y) = Q^j(\sigma_* y) = f(j, n) S^j(\sigma_* y). \end{aligned}$$

Thus, $f(j, n-1)(-1)^{p-1/2}\gamma(n) = f(j, n)$, where

$$\gamma(n) = (-1)^{(n-1)(p-1)/2} \cdot (p-1/2)! \bmod p.$$

A moment's calculation reveals that

$$f(j, n) \equiv (-1)^{n+j(p-1)/2+(n(n-1)/2)(p-1/2)} \cdot (p-1/2!)^{2j-n} \bmod p.$$

In terms of the operations $Q^{j}{}_{(p)}$, the Corollary 1 of Theorem 2. 3 becomes,

$$Q^{n}{}_{(p)}(x_r \cdot y_s) = \sum_{i=0}^{n} Q^{i}{}_{(p)}(x_r) Q^{n-o}{}_{(p)}(y_s).$$

The following is proved in [3], we give our proof for completeness.

THEOREM (Browder). *Let $\sigma_*: H_{r-1}(\Omega X; Z_p) \to H_r(X; Z_p)$ be the usual homology suspension. Let X be an $H^n{}_p$-space. Let $u \in H_{i-1}(\Omega X; Z_p)$, $v \in H_{j-1}(\Omega X; Z_p)$, then*

$$\sigma_* \psi_{n+1}(u, v) = (-1)^{i+n+1} \psi_n(\sigma_* u, \sigma_* v).$$

Proof. Let $\sigma_* u = x \in H_i(X; Z_p)$, $\sigma_* v = y \in H_j(X; Z_p)$. Also let $\bar{x} \in H_i(PX, \Omega X; Z_p)$, $\bar{y} \in H_j(PX, \Omega X; Z_p)$ be such that $\pi_*(\bar{x}) = x \in H_*(X, e; Z_p) = H(X; Z_p)$ and $\partial_*(\bar{x}) = u$, $\pi_*(\bar{y}) = y \in H_n(X, e; Z_p) = H(X; Z_p)$ and $\partial_*(\bar{y}) = v$. We must show there is a chain $c \in C_{i+j+n}(PX)$, such that: c is a relative cycle; i. e. $\{c\} \in H_{i+j+n}(PX, \Omega X; Z_p)$,

(1) $\pi_*\{c\} = (-1)^{i+n+1} \psi_n(x, y)$, and

(2) $\partial_*\{c\} = \psi_{n+1}(u, v)$.

For n even,

let c' be the chain in $W^{(n+1)} \otimes C(PX)^2$

$$c' = e_n \otimes \Delta(\bar{x} \otimes \bar{y}) - e_{n+1} \otimes (u \otimes \bar{y}) + (-1)^{ij} e_{n+1} \otimes (v \otimes \bar{x}).$$

Then $\partial c' = (-1)^{i-1} e_{n+1} \otimes \Gamma(u \otimes v)$.

For n odd,

$$\text{let } c' = e_n \otimes \Gamma(\bar{x} \otimes \bar{y}) + e_{n+1} \otimes (u \otimes \bar{y}) + (-1)^{ij} e_{n+1} \otimes (v \otimes \bar{x}).$$

Then $\partial c' = (-1)^{i} e_{n+1} \otimes \Delta(u \otimes v)$.

Now as in Theorem 3. 1, c' is contained in the special complex

$$[(W^{(n)}, e_{n+1}) \otimes C(\Omega X) \otimes C(PX)] + [W^{(n)} \otimes C(PX)^2]$$

which maps into $C(PX)$. By the same reasoning as in Theorem 3.1, the image c of $(-1)^{i+n+1}c'$ satisfies (1) and (2) above. The result follows.

Remark. From Theorem 1.2 it may be shown that certain "Adem Relations" hold for the homology operations in an iterated loop space. This may be done (following Adem) using the homology of the symmetric group. However, the device used by Adem to find the effect on odd classes from that on even classes does not apply here. Consequently, it appears necessary to compute the homology of the symmetric group with twisted coefficients to obtain the rest of the relations for the homology operations. (Note also that no universal space, such as a $K(\pi, n)$ exists in the case of homology operations.)

IV. Some theorems on Hopf algebras and spectral sequences. The object of this section is to present certain results, largely of an algebraic nature, which are useful in making explicit computations in Sevtion V. We use the notation and results of *The Structure of Hopf Algebras* by Milnor and Moore [11] and assume acquaintance with that paper on the part of the reader.

Our first theorem is a slight strengthening of the standard comparison theorem for spectral sequences essentially due to Zeeman [19].

THEOREM 4.1. *Suppose* $\{h^r\}\colon \{'E^r\} \to \{E^r\}$ *is a homomorphism of homology spectral sequences and*

$$'\chi_{a,b}\colon {}'E^2_{a,b} \to {}'E^2_{a,0} \otimes {}'E^2_{0,b} \text{ and } \chi_{a,b}\colon E^2_{a,b} \to E^2_{a,0} \otimes E^2_{0,b}$$

are isomorphisms such that the diagram

$$\begin{tikzcd}
{}'E^2_{a,b} \arrow[r, "'\chi_{a,b}"] \arrow[d, "h^2_{a,b}"] & {}'E^2_{a,0} \otimes {}'E^2_{0,b} \arrow[d, "h^2_{a,0} \otimes h^2_{0,b}"] \\
E^2_{a,b} \arrow[r, "\chi_{a,b}"] & E^2_{a,0} \otimes E^2_{0,b}
\end{tikzcd}$$

commutes

(α_n) $h^2_{a,0}$ *is an isomorphism for* $0 \leqq a \leqq n$,

and

(γ_n) $h^\infty_{a,b}$ *is an isomorphism if* $0 \leqq a + b \leqq n$ *and* $b \leqq n - 2$.

Then

(β_{n-2}) $h^2_{0,b}$ *is an isomorphism for* $0 \leqq b \leqq n - 2$, *and*

$h^r{}_{a,b}$ is an epimorphism if either $a+b \leqq n-1$, $a \leqq r-2$, and *$b \leqq n-2$ or $b \leqq n-r$ and $a \leqq n$,*

and

$h^r{}_{a,b}$ is a monomorphism if $b \leqq n-2$ and either $a+b \leqq n-1$ and *$n-(r-1) \leqq a$ or $a \leqq n-(r-1)$.*

Proof. (i) α_n and β_m for $m \leqq n-2$ imply

$(a_{r-1}) h^{r-1}{}_{a,b}$ is *epic* on $Z('E^{r-1}{}_{a,b})$ into $Z(E^{r-1}{}_{a,b})$

if either $a+b \leqq m+1$, $a \leqq r-2$, $b \leqq m$

or $b \leqq m-(r-2)$, $a \leqq n$

$(b_{r-1}) h^{r-1}{}_{a,b}$ is *monic* on $Z('^{r-1}{}_{a,b})$ and is an *isomorphism*

from $B('E^{r-1}{}_{a,b})$ onto $B(E^{r-1}{}_{a,b})$

if either $a+b \leqq n-1$, $n-(r-1) \leqq a$, $b \leqq m$

or $a \leqq n-(r-1)$, $b \leqq m$

$(c_r) h^r{}_{a,b}$ is *epic* if either $a+b \leqq m+1$, $a \leqq r-2$, $b \leqq m$
or $b \leqq m-(r-2)$, $a \leqq n$, and

$(d_r) h^r{}_{a,b}$ is *monic* if either $a+b \leqq n-1$, $n-(r-1) \leqq a$, $b \leqq m$
or $a \leqq n-(r-1)$, $b \leqq m$.

This argument is by induction on r. Clearly (c_r) and (d_r) are true for $r=2$.

It is easy to see that (a_{r-1}) implies (c_r) and (b_{r-1}) implies (d_r).

We now show that (c_r) and (d_r) imply (a_r). For $a+b \leqq m+1$, $a \leqq r-1$, $b \leqq m$, every element of $'E^r{}_{a,b}$ and $E^r{}_{a,b}$ is a cycle. Thus, (c_r) alone is sufficient to imply (a_r) in this case. For $\alpha \in Z(E^r{}_{a,b})$ with $b \leqq m-(r-1)$ and $a \leqq n$, by (c_r) there is an $'\alpha \in {}'E^r{}_{a,b}$ such that $h^r('\alpha) = \alpha$. We wish to show $'\alpha$ is a cycle. But $'d^r('\alpha) \in {}'E^r{}_{a-r,b+r-1}$ and $a-r \leqq n-r < n-(r-1)$ and $b+r-1 \leqq m$; and so, h^r is monic on $'d^r('\alpha)$. But $'d^r(h^{r\prime}\alpha) = 0$. Hence, $'d^r('\alpha) = 0$ and $a_r)$ is proved in this case also.

It is clear that (d_r) implies (b_r) for the cycles in $'E^r{}_{a,b}$ and also that h^r is monic on $B('E^r{}_{a,b})$ in the range stated. We need to show only that it is epic.

Let $\beta \in B(E^r{}_{a,b})$. There is a $\gamma \in E^r{}_{a+r,b-r+1}$ such that $d^r(\gamma) = \beta$.

If $a+b \leqq n-1$, $n-r \leqq a$ and $b \leqq m$, then $n-r+b \leqq n-1$ and $b-r+1 \leqq 0$. For $b-r+1<0$, $\gamma=0$ and so $\beta=0$. For $b-r+1=0$, $b=r-1$ and $a+r-1 \leqq n-1$; i.e., $a+r \leqq n$. But $n-r \leqq a$. Thus, $\gamma \in E^r{}_{n,0}$. Also, $b \leqq m$ and $b-r+1=0$ implies $0 \leqq m-r+1 \leqq m-(r-2)$. By (c_r) there is a $'\gamma \in {}'E^r{}_{n,0}$ such that $h^r('\gamma)=\gamma$. Hence, if $'\beta = {}'d^r\gamma \in {}'E^r{}_{a,b}$, then $h^r('\beta)=\beta$.

If $a \leqq n-r$ and $b \leqq m$, we have $\gamma \in E^r{}_{a+r,b-r+1}$ with $a+r \leqq n$ and $b-r+1 \leqq m-(r-1) \leqq m-(r-2)$. Thus, by (c_r) there is a $'\gamma \in {}'E^r{}_{a+r,b-r+1}$ such that $h^r('\gamma)=\gamma$; and so, $h^r('d^r{}'\gamma)=\beta$. Thus, h^r is epic as stated in (b_r).

(ii) α_n, γ_n, and β_m imply $h^{r+s}{}_{r,m-r+2}$ is an isomorphism for $r \geqq 2$ and $s \geqq 1$ if either

(a) $m+2 \leqq n-1$, $n-s \leqq 2r$ or

(b) $2r \leqq n-s$.

The argument is by induction downwards on $r+s$ and follows from the diagram

$$\begin{tikzcd}
 & & {}'E^{r+s}_{r,m-r+2} \arrow[d, "\cong"] & & \\
0 & {}'E^{r+s+1}_{r,m-r+2} \arrow[l] \arrow[d] & Z({}'E^{r+s}_{r,m-r+2}) \arrow[l] \arrow[d] & B({}'E^{r+s}_{r,m-r+2}) \arrow[l] \arrow[d] & 0 \arrow[l] \\
0 & {}'E^{r+s+1}_{r,m-r+2} \arrow[l] & Z({}'E^{r+s}_{r,m-r+2}) \arrow[l] & B({}'E^{r+s}_{r,m-r+2}) \arrow[l] & 0. \arrow[l] \\
 & & E^{r+s}_{r,m-r+2} \arrow[u, "\cong"'] & &
\end{tikzcd}$$

The isomorphism on the boundaries follows from (b_{r+s}) as

$$m+2=r+m-r+2 \leqq n-1, n-(r+s) \leqq r,\ m-r+2 \leqq m$$

in the case (a) and

$$r \leqq n-(r+s), m-r+2 \leqq m \text{ in the case (b).}$$

That the isomorphism holds for large $r+s$ is implied by (γ_n). The induction then follows by the "five lemma"

(iii) $h^r{}_{r,m-r+2}$ is an isomorphism on cycles for $r \geqq 2$ if either

(a) $m+2 \leqq n-1 \leqq 2r$

or

(b) $2r \leqq n-1$.

This follows from the diagram

$$\begin{tikzcd}
0 & {'E^{r+1}_{r,m-r+2}} \arrow[l] \arrow[d] & Z({'E^{r}_{r,m-r+2}}) \arrow[l] \arrow[d] & B({'E^{r}_{r,m-r+2}}) \arrow[l] \arrow[d] & 0 \arrow[l] \\
0 & E^{r+1}_{r,m-r+2} \arrow[l] & Z(E^{r}_{r,m-r+2}) \arrow[l] & B(E^{r}_{r,m-r+2}) \arrow[l] & 0. \arrow[l]
\end{tikzcd}$$

(iv) α_n, γ_n, β_m imply β_{m+1} if $m+2 \leqq n-1$.

That $h^r{}_{0,m+1}$ is an isomorphism, $r \geqq 2$, is shown by induction down on r with the following diagram

$$\begin{tikzcd}
0 & {'E^{0,m+1}} \arrow[l] \arrow[d] & {'E^{r}_{0,m+1}} \arrow[l] \arrow[d] & {'E^{r}_{r,m-r+2}} \arrow[l] \arrow[d] & Z({'E^{r}_{r,m-r+2}}) \arrow[l] \arrow[d] \\
0 & E^{r+1}_{0,m+1} \arrow[l] & E^{r}_{0,m+1} \arrow[l] & E^{r}_{r,m-r+2} \arrow[l] & Z(E^{r}_{r,m-r+2}). \arrow[l]
\end{tikzcd}$$

That $h^r{}_{r,m-r+2}$ is monic follows from (d_r) since $r+(m-r+2)=m+2 \leqq n-1$ and $m-r+2 \leqq m$. That it is epic follows from (c_r) since the only cases in which the groups are not zero are those in which $m-r+2 \geqq 0$;

Finally the second part of the conclusion of the theorem is simply the statements (c_r) and (d_r) with $m=n-2$.
i. e., $r \leqq m+2 \leqq n-1$. Of course, $m-r+2 \leqq m-(r-2)$.

The next theorem is from [11].

THEOREM 4.2. *If A is a connected Hopf algebra of finite type over a field of non-zero characteristic p, then A is primitively generated if and only if A is cocommutative and coassociative and for each positive dimensional element $u \in A^*$, $u^p = 0$.*

In our applications A will be the mod p homology algebra of an H-space; and so, the conditions "cocommutative" and "coassociative" will be automatically satisfied (as they are equivalent to commutativity and associativity of cup products in cohomolgy).

Recall (see [11]) that if B is a connected, commutative associative Hopf algebra and A is a sub-Hopf algebra, then letting $\bar{A}$ denote the module of positive dimensional elements of A and $C = B/\!/A = B/B\cdot\bar{A}$, C is a Hopf algebra and the projection map $\pi\colon B \to C$ is a map of Hopf algebras. Further, if there is a Hopf algebra map $j\colon C \to B$ such that $\pi \circ j$ is the identity, then $A \otimes C \cong B$ as Hopf algebras. Finally, recall also that if A is a sub-Hopf algebra of the connected commutative Hopf algebra B, then the diagram

$$\begin{tikzcd}
0 \arrow[r] & P(A) \arrow[r] \arrow[d] & P(B) \arrow[r] \arrow[d] & P(B/\!/A) \arrow[d] & \\
 & Q(A) \arrow[r] & Q(B) \arrow[r] & Q(B/\!/A) \arrow[r] & 0
\end{tikzcd}$$

is commutative and each of the horizontal rows is exact.

THEOREM 4.3. *Suppose B is a free, primitively generated, connected, commutative, associative Hopf algebra of finite type over a field of characteristic $p \neq 0$ and $f: A \to B$ is a map of Hopf algebras such that both f and $Q(f)$ are monomorphisms.*

Then A and $B/\!/A$ are primitively generated and

$$B \cong A \otimes (B/\!/A) \text{ as Hopf algebras.}$$

Proof. It is clear from the above diagram that $C = B/\!/A$ is primitively generated. Since $f^{*}: B^{*} \to A^{*}$ is an epimorphism, and B^{*} is commutative and associative and has all p-th powers zero (by Theorem 4.2), A^{*} has the same properties. Again by Theorem 4.2 it follows that A is primitively generated.

We shall define inductively direct sequences $\{A^{(n)}, f^{(n)}, g^{(n)}\}$ such that (1) $A^{(0)} = A$, (2)

$$\begin{tikzcd}
A^{(n)} \arrow[rr, "g^{(n)}"] \arrow[dr, "f^{(n)}"] & & A^{(n+1)} \arrow[dl, "f^{(n+1)}"] \\
& B &
\end{tikzcd}$$

is a commutative diagram of maps of Hopf algebras, (3) $A^{(n+1)} = A^{(n)} \otimes H(x^{(n+1)})$ as Hopf algebras, where $H(x^{(n+1)})$ is the free monogenic Hopf algebra generated by $x^{(n+1)}$, and (4) both $f^{(n)}$ and $Q(f^{(n)})$ are monomorphisms. Finally, we shall require that

(5) $$\varinjlim (f^{n}) : \varinjlim (A^{(n)}, g^{(n)}) \to B$$

be an isomorphism of Hopf algebras.

Before proceeding to the construction, let us observe that this is indeed sufficient to prove the theorem. It follows from (3) and (5) that $A^{(n+1)} \cong A^{(n)} \otimes (A^{(n+1)}/\!/A^{(n)})$ as Hopf algebras, that

$$A^{(n+1)} \cong A \otimes (A^{(1)}/\!/A) \otimes \cdots \otimes (A^{(n+1)}/\!/A^{(n)})$$

as Hopf algebras, and finally that $B \cong A \otimes \{A^{(1)}/\!/A \otimes A^{(2)}/\!/A^{(1)} \otimes \cdots\}$ as Hopf algebras. Letting $C' = \{A^{(1)}/\!/A \otimes A^{(2)}/\!/A^{(1)} \otimes \cdots\}$, we then have

$$C = B/\!/A \cong C'; \text{ and so,}$$

$$B \cong A \otimes C \text{ as Hopf algebras.}$$

Assume $g^{(i-1)}$, $f^{(i)}$, $A^{(i)}$ defined and having the stated properties for $i \leqq n$. If there is an element of $Q(B)$ not in the image of $f^{(n)}$ (if not the induction is complete), let $x^{(n+1)}$ be a primitive element of B projecting onto one of least dimension. Let $A^{(n+1)} = A^{(n)} \otimes H(x^{(n+1)})$, let $g^{(n)}$ be the inclusion $A^{(n)} \rightarrow A^{(n)} \otimes 1 \rightarrow A^{(n+1)}$, and let $f^{(n+1)}$ be the extension of $f^{(n)}$ defined by sending $x^{(n+1)}$ into $x^{(n+1)}$. Clearly (1), (2) and (3) are true. To show that $f^{(n+1)}$ is a monomorphism, it is sufficient to show that $P(f^{(n+1)})$ is. As $P(f^{(n)})$ is a monomorphism, it is thus sufficient to show that no x^{p^s} (letting $x = x^{(n+1)}$) lies in the image of $f^{(n)}$. Suppose x^{p^s} is in the image of $f^{(n)}$. As x^{p^s} is primitive and decomposable and $Q(f^{(n)})$ is a monomorphism, there is an element $a \in A^{(n)}$ such that

$$x^{p^s} = f^{(n)}(a^p).$$

Thus, $(x^{p^{s-1}} - f^{(n)}(a))^p = 0$, and since B is free, $x^{p^{s-1}} = f^{(n)}(a)$. Since x is not in the image of $f^{(n)}$, this leads to a contradiction. To see that $Q(f^{(n+1)})$ is a monomorphism, it is sufficient to abserve thot

$$\dim Q(A^{(n+1)}) = \dim Q(A^{(n)}) + 1$$

and that by construction

$$\dim f^{(n+1)} Q(A^{(n+1)}) = \dim f^{(n)} Q(A^{(n)}) + 1.$$

But by assumption $Q(f^{(n)})$ is a monomorphism. Thus, (4) is proved. The statement (5) follows from the fact that B is of finite type and $f^{(n)}$ is a monomorphism together with the method of choosing $x^{(n+1)}$ (by exhaustion!).

A commutative Hopf algebra homology spectral sequence $\{E^r, d^r\}$ (over Z_p) is a homology spectral sequence in which each E^r is a commutative Hopf algebra over Z_p, d^r is a derivative, and $E^{r+1} \cong H_*(E^r)$ as Hopf algebras.

Definition. A commutative Hopf algebra homology spectral sequence $\{E^r, d^r\}$ over Z_p is a *model spectral sequence* provided either

(a) $E^2{}_{*,0} \cong E(x, n)$ and $E^2{}_{0,*} \cong P(y, n-1)$, these isomorphisms being as Hopf algebras, where n is odd and $d^n(x \otimes y^i) = y^{i+1}$ for all $i \geqq 0$, ($E(x, n)$ denotes the monogenic exterior Hopf algebra with generator x in dimension x and $P(y, n-1)$ the monogenic polynomial Hopf algebra with generator y in dimension $n-1$) or

(b) $E^2{}_{*,0} \cong P(x, n)$ and $E^2{}_{0,*} \cong \bigotimes_{k \geqq 0} E(y_k, p^k n - 1) \otimes \bigotimes_{j > 0} P(z_j, p^j n - 2)$, these isomorphisms also being as Hopf algebras, where n is even and

$$d^{np^k}(x^{p^k}) = y_k \text{ for } k \geqq 0 \text{ and}$$

$$d^{np^j(p-1)}((x^{p^j})^{p-1} \otimes y_j) = z_{j+1} \text{ for } j \geqq 0.$$

(The spectral sequence (a) is that of the mod p homology ($p \neq z$) of the fibration $\Omega S_n \to PS_n \to S_n$, n odd, and (b) is the spectral sequence of the mod p homology of the fibration $\Omega^2 S_n \to P\Omega S_n \to \Omega S_n$, n odd (see Moore [12]). In the latter case it is also true that $\partial_{*p}(y_k) = + z_k$, where ∂_{*p} denotes the homology Bockstein of the sequence $0 \to Z_p \to Z_{p^2} \to Z_p \to 0$.)

It is known [11] that if A is a connected, free, associative, commutative, primitively generated mod p Hopf algebra of finite type, then A is isomorphic, as Hopf algebra, to a tensor product of free monogenic Hopf algebras; i.e., to a tensor product of exterior and polynomial Hopf algebras each having one generator. Associated with each of these free monogenic Hopf algebras is a model spectral sequence with that Hopf algebra as the term $E^2_{*,0}$. The tensor product of these model spectral sequences, one for each monogenic factor in the decomposition of A, is a *canonical spectral sequence for* A. It will be shown later in this section that two canonical spectral sequences for A corresponding to different decompositions of A into monogenic factors are isomorphic as spectral sequences of Hopf algebras. We shall denote such a spectral sequence by $\{C^r(A)\}$.

THEOREM 4.4. *Suppose* $\{E^r, d^r\}$ *is an associative, commutative Hopf algebra homology spectral sequence over* Z_p, $p \neq 2$, *such that* $E^\infty = 0$ *and if* $x \in E^2_{2n,0}$ *such that* $d^r x = 0$ *for all* $r < 2n$, *then*

(i) $d^r(x^p) = 0$ *for all* $r < 2np$,

(ii) $d^r(x^{p-1} \otimes \tau(x)) = 0$ *for all* $r < 2n(p-1)$, *and*

(iii) *if* $d^{2n(p-1)}(x^{p-1} \otimes \tau(x)) \neq 0$, *then* $d^{2np}(x^p) \neq 0$.

Then

(a) *if* $E^2_{*,0}$ *is transgressively generated, it is primitively generated and free,*

(b) *if* $E^2_{*,0}$ *is primitively generated and free, and*

$$\bigotimes_{i \in I} E(x_i, n_i) \otimes \bigotimes_{j \in J} P(y_j, m_j)$$

is a Hopf algebra decomposition of $E^2_{*,0}$ *into its monogenic parts, and there do not exist* n_i, m_j, m_k *and integers* f *and* g, $f \geqq 1$ *and* $g \geqq 1$, *such that either*

(A) $m_j = 2 + (n_i - 1) \cdot p^f$, *or*

(B) $m_j = 2 + (m_k \cdot p^g - 2) \cdot p^f$,

then $E^2_{*,0}$ *is trangressively generated, and*

(c) *if* $E^2_{*,0}$ *is transgressively generated, then* $\{E^r\}$ *is isomorphic to* $\{C^r(E^2_{*,0})\}$ *as spectral sequences of algebras.*

In proving part (b) of this theorem we shall need

LEMMA 4.5. *Suppose* A *and* B *are connected, associative, commutative, Hopf algebras of finite type,* A *is free,* n *is a positive integer, and* $f\colon A \to B$ *is an algebra homomorphism which is an isomorphism in dimensions less than* n. *Then if* $Q(A)_n = 0$ *and* $n \neq p^k \cdot m$ *for some integer* k *and even integer* m *such that* $Q(A)_m \neq 0$, *then* f *is a monomorphism in dimension* n.

Proof of Lemma. As f is an algebra isomorphism in dimensions less than n, $Q(f)$ is an isomorphism in dimensions less than n and, in particular, $\dim Q(A)_i = \dim Q(B)_i$ for $i < n$. As an algebra $B \cong \otimes E(\zeta_i) \otimes P^{f_j}(\eta_j)$, where $P^{f_j}(\eta_j)$ denotes the truncated polynomial algebra in η_j of height p^{f_j}. Further as B is free in dimensions less than n, $(\dim \bar{\eta}_j) \cdot p^{f_j} > n$. Let $\bar{B} \equiv \otimes E(\bar{\zeta}_i) \otimes \bigotimes P(\bar{\eta}_j)$ and $\pi\colon \bar{B} \to B$ be the unique algebra extension of $\pi(\bar{\zeta}_i) = \zeta_i$ and $\pi(\bar{\eta}_j) = \eta_j$, where the tensor product is over just those $\bar{\zeta}_i$ and $\bar{\eta}_j$ of dimension less than n. The map π is a map of algebras, is an isomorphism in dimensions less than n, and is a monomorphism in dimension n. There is a sub-Hopf algebra $\bar{A} = \bigotimes E(x_i) \otimes \bigotimes P(y_j)$ of A such that $Q(\bar{A})_i = 0$ for $i \geqq n$ and the inclusion map $\iota\colon \bar{A} \to A$ is an isomorphism in dimensions less than or equal to n. Define $\bar{f}\colon \bar{A} \to \bar{B}$ to be the unique algebra extension of $\bar{f}(x) = \pi^{-1} f \iota(x)$ for each $x = x_i$ or $x = y_j$. Then

$$\begin{tikzcd}
\bar{A} \arrow[r, "f"] \arrow[d, "\iota"'] & \bar{B} \arrow[d, "\pi"] \\
A \arrow[r, "f"] & B
\end{tikzcd}$$

is a commutative diagram of algebra homomorphisms; and so, $Q(\pi \circ \bar{f}) = Q(f \circ \iota)$. In dimension $i < n$,

$$\dim Q(f \circ \iota)\,(Q(\bar{A})_i) = \dim Q(\bar{A})_i = \dim Q(A)_i \text{ and}$$
$$\dim Q(\pi)\,(Q(\bar{B})_i) = \dim Q(\bar{B})_i = \dim Q(B)_i.$$

Were $Q(\bar{f})_i$ not an epimorphism, then

$$\dim Q(\pi \circ \bar{f})\,(Q(\bar{A})_i) < \dim Q(\bar{B}_i) = \dim Q(f \circ \iota)\,(Q(\bar{A})_i).$$

Thus, $Q(\bar{f})$ is an epimorphism; and since $\bar{f}$ is a map of algebras, $\bar{f}$ is an epimorphism. A counting argument then shows that $\bar{f}$ is a monomorphism, and the conclusion follows immediately.

Proof of Theorem 4.4.

(a) As any transgressive element in a Hopf algebra spectral sequence is primitive, it is immediate that $E^2{}_{*,0}$ is primitively generated. To show it is free, it will suffice to show that if x is an even dimensional generator in $E^2{}_{*,0}$, then $x^{p^i} \neq 0$ for all i. Suppose $x^{p^f} \neq 0$ and $x^{p^{f+1}} = 0$. Let $y = x^{p^f}$. Both y and y^p $(= 0)$ transgress, by hypothesis. Let $w = \tau(y)$ and let $\bar{w}$ be an element in $E^{0,*}{}_2$, the dual cohomology spectral sequence, such that $\langle \bar{w}, w \rangle \neq 0$. As $\langle \bar{w}, d^n y \rangle \neq 0$, $\langle \delta_n \bar{w}, y \rangle \neq 0$ and $\bar{y} = \delta_n \bar{w} \neq 0$, where $n = \dim y$. Since $y^p = 0$, by hypotheses (ii) and (iii), all d^r on $y^{p-1} \otimes w$ are 0. Thus, $\bar{y}^{p-1} \otimes \bar{w}$ is not a coboundary in E_r. Since $E^2{}_{*,0}$ is primitively generated, by Theorem 4.2 we have $\bar{y}^p = 0$. But $\delta_i(\bar{y}^{p-1} \otimes \bar{w}) = 0$ for $i < n$ and $\delta_n(\bar{y}^{p-1} \otimes \bar{w}) = \bar{y}^p = 0$. Thus, all δ_r on $\bar{y}^{p-1} \otimes \bar{w}$ are 0; and so, $y^{p-1} \otimes w$ is not a boundary in any E^r. The class $y^{p-1} \otimes w$ thus leads to a non-zero class in E^∞, which is a contradiction.

(b) Let $\mathscr{X}$ be the set of all generators of $E^2{}_{*,0}$ which transgress. $\mathscr{X} \neq \phi$, as the least dimensional generator of $E^2{}_{*,0}$ transgresses.

Suppose $\mathscr{X}$ does not generate $E^2{}_{*,0}$. Let z be a least dimensional generator of $E^2{}_{*,0}$ not in the algebra generated by $\mathscr{X}$. Among all generators of the dimension of z not in that algebra let ν be one which is maximally transgressive; i. e.,

$$d^s \nu = 0, \ s < r$$

$$\mu = d^r \nu \neq 0, \text{ and}$$

if $\bar{\nu}$ is any other generator of the dimension of z not in the algebra generated by $\mathscr{X}$ and

$$d^t \bar{\nu} = 0, \ t < \bar{r},$$

then $\bar{r} \leqq r$. Further, if there is such an element ν which is primitive, choose it. Let $n = \dim z = \dim \nu$.

For each $x \in \mathscr{X}$, let $\{C^r(x)\}$ be the model spectral sequence for $E(x)$ or $P(x)$, as the case may be. It follows as a routine verification from the hypothesis (i), (ii) and (iii) that there is an algebra homomorphism of spectral sequences

$$\{h^r(x)\} : \{C^r(x)\} \to \{E^r\}.$$

Let $'E = \bigotimes_{x \in \mathscr{X}} C(x)$ and

$$\{h^r\} : \{'E^r\} \to \{E^r\}$$

be the map induced by the $\{h^r(x)\}$. $\{h^r\}$ is an algebra homomorphism of spectral sequences and $h^2{}_{i,0}$ is an isomorphism for $i \leqq n-1$.

Let μ_2 be an element of $E^2{}_{n-r,r-1}$ projecting onto μ. There is an element

$'\mu_2 \in {}'E^2{}_{n-r,r-1}$ such that $h^2('\mu_2) = \mu_2$. Since μ_2 projects onto μ, the class of μ_2 in E^s, $s < r$, is not a boundary under d^s. Thus, the same is true for $'\mu_2$. Since $h^s{}_{a,b}$ is a onomorphism for $a + b \leqq n - 2$ and $b \leqq n - 3$, no $'d^s$ on the class of $'\mu_2$ is non-trivial for $s < r$, unless possibly some $'d^s('\mu_2) \in {}'E^s{}_{0,n-2}$. We shall return in the next paragraph to show this cannot occur. Thus, $'\mu_2$ projects onto a class $'\mu$ such that $h^r('\mu) = \mu$. If there were no element $'\gamma \in {}'E^2{}_{n,0}$ such that $'d^r('\gamma) = {}'\mu$, then $'\mu$ would give rise to a non-zero term of $'E^\infty{}_{n-r,r-1}$. But $'E^\infty = 0$. Thus, there is a $'\gamma \in {}'E^2{}_{n,0}$ such that $'d^r('\gamma) = {}'\mu$.

Hence,

$$d^r\{\nu - h^r('\gamma)\} = \mu - h^r('\mu) = 0,$$

and $\nu - h^2('\gamma)$ is more transgressive than ν. As $h^2('\gamma)$ is in the algebra generated by $\mathfrak{X}$, $\nu - h^2('\gamma)$ is a generator of $E^2{}_{*,0}$ not in that algebra; and so, the assumption that $\mathfrak{X}$ does not generate $E^2{}_{*,0}$ leads to a contradiction; i. e., $E^2{}_{*,0}$ is transgressively generated.

We have only to show that it is impossible for $'\lambda_{n-r} = {}'d^{n-r}('\mu_2)$ to be non-zero. Since $h^2{}_{i,0}$ is an isomorphism for $i \leqq n - 1$ and h^∞ is an isomorphism (both $'E^\infty$ and E^∞ are trivial), $h^{n-r}{}_{0,i}$ is an isomorphism for $i \leqq n - 3$. If $'\lambda_{n-r} \in {}'E^{n-r}{}_{0,n-2}$ were non-zero, then since $h^{n-r}('\lambda_{n-r}) = 0$, by Lemma 4. 5 either it is indecomposable in $'E^{n-r}{}_{0,*}$ or $n - 2 = p^k \cdot m$, where m is the dimension of some even dimensional generator of $'E^{n-r}{}_{0,*}$. If $'\lambda_{n-r}$ were indecomposable, then since $'E^2{}_{0,*} \rightarrow {}'E^{n-r}{}_{0,*}$ is an epimorphism of algebras, there is a generator $'\lambda_2 \in {}'E^2{}_{0,*}$ which projects onto $'\lambda_{n-r}$. In the canonical spectral sequence $'E$ any odd dimensional generator in $'E^2{}_{0,*}$ is the transgression of an element in $'E^2{}_{*,0}$ and thus cannot be the image under $'d^{n-r}$ of some element. The even dimensional generators are either transgressions of odd dimensional generators in $'E^2{}_{*,0}$, and cannot be in the image of $'d^{n-r}$, or are the images of elements of the form $('x^{p-1} \otimes \tau(x'))$, where $'x = {}'z^{p^i}$ for some even dimensional generator $'z$. Let $m = \dim {}'x$ and $'y = \tau('x)$. Let x and y denote the images of $'z$. Let $m = \dim {}'x$ and $'y = \tau('x)$. Let x and y denote the images of $'x$ and $'y$ under h. In E^m both x and y are primitive and $(x^{p-1} \otimes y)$ is *not primitive*. For $'\lambda_{n-r}$ to be the differential of $('x^{p-1} \otimes \tau('x))$ we must have $\mu = (x^{p-1} \otimes y)$ and $d^r\nu = \mu$, $r = m$. Since ν is a generator of $E^2{}_{*,0}$, it is also one in $E^m{}_{*,0}$. As $E^m{}_{*,0}$ is a sub-Hopf algebra of $E^2{}_{*,0}$, it is primitively generated by the first part of Theorem 4. 3. Thus, $\nu = w + \gamma$ in $E^m{}_{*,0}$, where w is a primitive generator in $E^m{}_{*,0}$ and γ is decomposable. Since ν is a generator in $E^2{}_{*,0}$ and γ is decomposable, w is a primitive generator in $E^2{}_{*,0}$. It is then at least as transgressive as ν; and so, by our initial assumption on ν, ν is primitive. It is impossible for $d^r\nu$ to be μ, which is not primitive in E^m. The only case remaining is for $n-2=p^k\cdot m$, where m is the dimension of some even dimensional generator of $'E^{n-r}_{0,*}$, $k>0$. Such a generator is the image of a generator y in $'E^2_{0,*}$. Such generators arise in one of two ways:

(1) $y=\tau(x)$, where x is an odd dimensional generator of $'E^2_{*,0}$, or

(2) $y=\tau(w)$, where $w=(x^{p^f})^{p-1}\otimes\tau(x^{p^f})$ and x is an even dimensional generator of $'E^2_{*,0}$.

In case (1), $n-2=p^k\cdot\dim y$; i.e., $n=2+(\dim x-1)\cdot p^k$, which is excluded by hypothesis. In case (2), $'\lambda=\tau(w\otimes y^{p^k-1})$; and so, $'\mu=w\otimes y^{p^k-1}$. If $k\geqq 1$, then

$$\begin{aligned}\dim\nu&=1+\dim w+\dim(y^{p^k-1})\\&=2+p^k\cdot\{mp^{f+1}-2\},\text{ where}\end{aligned}$$

$m=\dim x$ and $f\geqq 0$. This is also excluded by hypothesis.

(c) One picks a multiplicative base of transgressive generators and for each of them maps a model spectral sequence into $\{E^r\}$. The tensor product of these maps (as in part (b)) is an algebra map of spectral sequences and induces isomorphisms on $E^2_{*,0}$ and E^∞. By Theorem 4.1 the conclusion follows.

In view of the complicated technical nature of the hypotheses, it is appropriate to discuss them. The standing hypotheses (i), (ii) and (iii) are vacuously true if $E^2_{*,0}$ is an exterior Hopf algebra.

One can see that conditions (i) and (ii) hold in a loop space $X=\Omega Y$, by considering the canonical map $s\Omega X\to X$, which in homology is onto all transgressive elements. Also, since $\Omega Y\to\Omega s\Omega Y\to\Omega Y$ is a retraction, the diagram

$$\begin{tikzcd}\Omega s^2\Omega X \arrow[r] & \Omega sX \arrow[r] & X\\ & X \arrow[u] & \end{tikzcd}$$

shows that if conditions (i) and (ii) hold in mod p homology spectral sequence of the fibration

$$\Omega^2s^2W\to P\Omega s^2W\to\Omega s^2W$$

for every space W, then they hold for the fibration

$$\Omega X\to PX\to X$$

for every loop space X. Browder, in his thesis, has determined explicitly the structure of the mod p homology spectral of the fibration

$$\Omega^2s^2W\to P\Omega s^2W\to\Omega s^2W$$

and the conditions (i) and (ii) hold. He has asserted elsewhere that condition (iii) also is valid. We shall give later (Theorem 4.7) a homology analogue of Kudo's Transgression Theorem which gives more precise information about the conditions (i), (ii) and (iii) in case X is a special $H^{(p-2)}{}_p$-space. Regarding the hypothesis in (b), there are examples of non-transgressively generated spectral sequences satisfying the other hypotheses which do not satisy (A) and also examples which do satisfy (B). Finally an inspection of the proof of (b) shows that the only place in which the Hopf algebra structure is used is in showing $h^s{}_{0,n-2}$ is a monomorphism, where n is the dimension of some generator in $E^2{}_{*,0}$. If $E^2{}_{*,0}$ is an exterior algebra, in each of the canonical spectral sequences, $'E^2{}_{0,*}$ is a polynomial algebra and has no odd dimensional elements. Thus, $h^s{}_{0,n-2}$ is vacuously a monomorphism. Hence, we have the following dual of Borel's Transgression Theorem:

THEOREM 4.6. *Suppose $\{E^r, d^r\}$ is an associative, commutative, algebra homology spectral sequence over Z_p such that $E^\infty = 0$ and $E^2{}_{*,0}$ is an exterior algebra of finite type. Then each generator of $E^2{}_{*,0}$ is transgressive and $\{E^r\} \cong \{C^r(E^2{}_{*,0})\}$ as algebra spectral sequences. In particular, $E^2{}_{0,*}$ is the polynomial alegebra generated by the transgressions of the generators of $E^2{}_{*,0}$.*

As for the standing hypotheses (i), (ii), and (iii) of Theorem 4.4, we have

THEOREM 4.7. *If X is a special $H_p{}^{(p-2)}$-space, p odd,* (see Theorem 1.1), *$\{E^r\}$ is the mod p homology spectral sequence of the fibration $\Omega X \to PX \to X$, and $x \in E^2{}_{2n,0}$ is a transgressive class and $y \in E^2{}_{0,2n-1}$ is a class such that $\tau(x) = y$ in E^{2n}, then*

$$\tau(x^p) = Q^n{}_{(p)}(y) \text{ in } E^{2np}$$

and

$$\tau(x^{p-1} \otimes y) = + \partial^*{}_p Q^n{}_{(p)}(y) \text{ in } E^{2n(p-1)},$$

where $\partial^*{}_p$ is the homology Bockstein homomorphism associated with the exact coefficient sequence

$$0 \to Z_p \to Z_{p^2} \to Z_p \to 0.$$

Proof. Since $\sigma_*(y) = x$, it follows from Theorem 3.1 that $\sigma_* Q^n{}_{(p)}(y) = Q^n{}_{(p)}(x)$, but $Q^n{}_{(p)}(x) = x^p$. Thus, x^p is transgressive and $\tau(x^p) = Q^n{}_{(p)}(y)$ in E^{2np}. More specifically, the class c'' for $i = 0$ in the proof of Theorem 3.1 is relevant.

$$c'' = e_{p-1} \otimes \epsilon_{(2n-1)p+1} + e_{p-2} \otimes \Delta^{p-2} \epsilon_{(2n-1)p+2} + \cdots + (-1)^{p-1/2} e_0 \otimes \epsilon_{2np}.$$

$$\partial c'' = e_{p-1} \otimes \epsilon_{(2n-1)p} + (\Gamma - \Delta^{p-1}) \{e_{p-2} \otimes \epsilon_{(2n-1)p+1} - e_{p-3} \otimes \epsilon_{(2n-1)p+2}$$
$$+ \cdots + (-1)^{p-1/2} e_0 \otimes \epsilon_{(2np)-1}\}.$$

The polynomial $\Gamma - \Delta^{p-1}$ is congruent to zero mod p; i.e., $\Gamma - \Delta^{p-1} = pR(\alpha)$. This implies that $R(1) = 1$; and so, the sum of the coefficients of $R(\alpha)$ is 1. Under the map ϕ of the proof of Theorem 3.1, we have

$$\begin{aligned}
&\phi(\epsilon_{(2n-1)p}) = (\partial z)^p, \\
&\phi(\epsilon_{(2n-1)p+1}) = z \otimes (\partial z)^{p-1}, \\
&\quad\vdots \\
&\phi(\epsilon_{2np-1}) = P(\alpha) \cdot z^{p-1} \otimes (\partial z), \text{ and} \\
&\phi(\epsilon_{2np}) = \gamma \cdot z^p,
\end{aligned}$$

where $z \in C(PX)$ such that $\partial z = y + p \cdot C$, $y \in C(\Omega X)$, $\pi_{\ddagger}(z) = \{x\} \bmod p$, and $P(\alpha)$ is some polynomial in α. We can determine $P(\alpha)$ as follows:

$$\begin{aligned}
\Gamma \cdot P(\alpha) \cdot z^{p-1} \otimes (\partial z) &= \phi(\Gamma \epsilon_{2np-1}) = \phi(\partial \epsilon_{2np}) \\
&= \partial \phi(\epsilon_{2np}) = \gamma \cdot \partial z^p = \gamma \cdot \Gamma \cdot z^{p-1} \otimes (\partial z).
\end{aligned}$$

Thus, $\Gamma \cdot P(\alpha) = \gamma \cdot \Gamma$ and so $P(\alpha) = \gamma + \Delta \cdot Q(\alpha)$. (Actually, $Q(\alpha)$ is a polynomial in Δ, but this will not be needed.) Letting c denote the image of c'' in $C(PX)$, we have

$$\begin{aligned}
&c = a_1 + \cdots + a_p \text{ and} \\
&\partial c = b_0 + p \cdot R(\alpha)\{b_1 - b_2 + \cdots + (-1)^{p-1/2} b_{p-1}\},
\end{aligned}$$

where $a_p = (-1)^{p-1/2} \cdot \gamma \cdot z^p$, $b_0 = (-1)^{p-1/2} \cdot \gamma \cdot Q^{(n)}{}_p(y) \bmod p$ and $\bmod p^2$ (since $(\partial z)^p = (y + pc)^p = y^p \bmod p^2$) and

$$p \cdot b_{p-1} = p \cdot \{\gamma + \Delta \cdot Q(\alpha)\}(z^{p-1} \cdot (y + pc)) = p\{\gamma + \Delta \cdot Q(\alpha)\}(z^{p-1} \cdot y) \bmod p^2.$$

Thus, modulo p^2 we have

$$\begin{aligned}
&(-1)^{p-1/2} \cdot \gamma \cdot Q^{(n)}{}_p(y) \\
&\qquad = \partial c - R(\alpha) \cdot \{pb_1 - pb_2 + \cdots + (-1)^{p-1/2} p \cdot [\gamma + \Delta Q(\alpha)](z^{p-1} \cdot y).
\end{aligned}$$

Since Δe_0 is a boundary, $\Delta(z^{p-1} \cdot y)$ is a boundary and since PX is homotopy commutative and the dimension of z is even, for any i we have that $(\alpha^i - 1)(z^{p-1} \cdot y)$ is a boundary. Since the sum of the coefficients in $R(\alpha)$ is 1, it follows that $(R(\alpha) - 1)(z^{p-1} \cdot y)$ is a boundary.. Hence, we have

$$\begin{aligned}
&(-1)^{p-1/2} \cdot \gamma \partial^*{}_p Q^{(n)}{}_p(y) \\
&\qquad = \partial d - (-1)^{p-1/2} p \cdot \gamma \cdot (z^{p-1} \cdot y) - R(\alpha) \cdot \{pb_1 - pb_2 + \cdots + pb_{p-2}\}.
\end{aligned}$$

mod p^2; and so,

$$(-1)^{p-1/2}\cdot\gamma\partial\div_p Q^{(n)}{}_p(y)$$
$$=+\gamma\cdot(-1)^{p-1/2}(z^{p-1}\cdot y)+R(\alpha)\cdot\{b_1-b_2+\cdots+b_{p-2}\}.$$

The filter degrees of the terms b_1 to b_{p-2} are such that they all give zero in $E^2{}_{2n(p-1),2n-1}$. Hence,

$$\partial*_p Q^{(n)}{}_p(y)=+\tau(z^{p-1}\cdot y).$$

We next show that a canonical spectral sequence for A is independent of the particular monogenic decomposition of A used in its construction. To this end the follownig lemma suffices:

LEMMA 4.8. *If* $\{'E^r\}$ *and* $\{E^r\}$ *are canonical spectral sequences* (*mod* p, $p\neq 2$) *and* $f\colon {}'E^2{}_{*,0}\to E^2{}_{*,0}$ *is a Hopf algebra homomorphism, then there is a Hopf algebra spectral sequence homomorphism* $\{f^r\}\colon\{'^r\}\to\{E^r\}$ *such that* $f^2{}_{*,0}=f$.

Proof. It suffices to prove this for the case where $'E^2{}_{*,0}$ is a monogenic Hopf algebra, $H(u)$. $f(u)$, being a primitive element of $E^2{}_{*,0}$, is either $\sum a_i v_i$, where the v_i are odd dimensional generators of the particular decomposition of $E^2{}_{*,0}$ used in describing $\{E^r\}$ or $\sum b_i w_i{}^{p^{k_i}}$, where the w_i are even dimensional generators. In the former case, each of the v_i transgresses into a primitive element $y_i\in E^2{}_{0,*}$. We let $f^2{}_{0,*}('\tau(u))=\sum a_i y_i$. This induces a Hopf algebra map $f^2\colon{}'E^2\to E^2$. There is only one non-trivial differential d^r, $r=\dim u$, and the induced f^r commutes with it. Hence, f^r is defined as a Hopf algebra homomorphism of spectral sequences. The latter case is slightly more troublesome. Let $z_i=b_i w_i{}^{k_i}$. Then z_i is primitive and transgresses into the primitive element t_i. Let $f^2{}_{0,*}('\tau(u))=\sum t_i$. As $(\sum z_i)^p=\sum(z_i)^p$, we let $f^2{}_{0,*}('\tau(u^p))=\sum\tau(z_i)^p$. We must show that $(\sum z_i)^{p-1}\otimes(\sum t_i)$ transgresses. The terms $z_{i_1}{}^{j_1}\cdots z_{i_k}{}^{j_k}\cdot t_l$ with at least two non-zero j_s's as exponents are boundaries under d^r, $r=\dim u$, of $z_{i_1}{}^{j_1}\cdots z_{i_k}{}^{j_k}\cdot z'_l$ since $j_1+\cdots+j_k+1=p$, two j_s's are non-zero and so no z_s has exponent p. The terms $z^{p-1}{}_i\otimes t_j$ for $i\neq j$ are boundaries of $z^{p-1}{}_i\cdot z_j$. Thus, in E^r, $(\sum z_i)^{p-1}\otimes(\sum t_i)-(\sum z^{p-1}{}_i\otimes t_i)$ is a boundary. As $(\sum z^{p-1}{}_i\otimes t_i)$ transgresses, in fact to a primitive class, it follows that $(\sum z_i)^{p-1}\otimes(\sum t_i)$ also does. Let

$$f^2{}_{0,*}('\tau(u^{p-1}\otimes{}'\tau(u)))=\tau(\{f^2{}_{*,0}(u)\}^{p-1}\otimes f^{\star}{}_{*,0}{}'\tau(u)).$$

Clearly, a similar argument is valid for terms of the form u^{p^i} and $(u^{p^i})^{p-1}\otimes{}'\tau(u^{p^i})$. Thus, there is described a Hopf algebra map $f^2\colon{}'E^2\to E^2$ which commutes with differentials.

In particular, we have that if $\{'C^r(A)\}$ and $\{C^r(A)\}$ are canonical spectral sequences induced by different monogenic decompositions of A, then the extension of the identity map of A into itself is a Hopf algebra homomorphism of spectral sequences, and since it induces isomorphisms on $'C^2_{*,0}(A)$ and $'C^\infty$ levels, it is then a Hopf algebra isomorphism of spectral sequences.

THEOREM 4.9. *Suppose* $\{Y^r\}$ *is a canonical spectral sequence,* $\{X^r\}$ *is a spectral sequence of algebras,* $X^2_{*,0}$ *is a Hopf algebra, and* $\{f^r\}: \{X^r\} \to \{Y^r\}$ *is an algebra homomorphism of spectral sequences such that* $f^2_{*,0}$ *and* $Q(f^2_{*,0})$ *are monomorphism,* $f^2_{*,0}$ *is a map of Hopf algebras and* $X^\infty = 0$. *Then*

(a) $\{X^r\}$ *is isomorphic as algebra spectral sequence to a canonical spectral sequence*

(b) $f^2_{0,*}$ *and* $Q(f^2_{0,*})$ *are monomorphisms*

(c) *if* $\{X^r\}$ *is a spectral sequence of Hopf algebras and* $\{f^r\}$ *is a Hopf algebra homomorphism of spectral sequences, then* $\{X^r\}$ *is isomorphic as Hopf algebra spectral sequence to a canonical spectral sequence.*

Proof. By Theorem 4.3 there is a Hopf algebra D and a homomorphism $g: D \to Y^2_{*,0}$ such that $X^2_{*,0} \otimes D \xrightarrow{(f^2_{*,0} \otimes g)} Y^2_{*,0}$ is a Hopf algebra isomorphism. By Lemma 4.8 there is an extension

$$\{(fg)^r\}: \{C^r(X^2_{*,0} \otimes D)\} \to \{Y^r\}.$$

$\{(fg)^r\}$ is a Hopf algebra isomorphism of spectral sequences. But there is an isomorphism

$$\{h^r\}: \{C^r(X^2_{*,0})\} \otimes \{C^r(D)\} \to \{C^r(X^2_{*,0} \otimes D)\}$$

as Hopf algebra spectral sequences.

Thus, we have the sequence of maps

$$\{X^r\} \xrightarrow{\{f^r\}} \{Y^r\} \xrightarrow{\{k^r\}} \{C^r(X^2_{*,0})\} \otimes \{C^r(D)\} \xrightarrow{\{l^r\}} \{C^r(X^2_{*,0})\}$$

in which $\{k^r\}$ and $\{l^r\}$ are Hopf algebra maps and $l^2_{*,0} \circ k^2_{*,0} \circ f^2_{*,0}$ is the identity.

The conclusions (a) and (c) are then immediate from Theorem 4.1; (b) is clear since Q(identity) is the identity on Q.

V. Homology of iterated loop spaces.

Definition 5.1. A *spectrum* $S = \{X_i, f_i\}$ (see [18]) is a sequence of spaces $\{X_i\}$, $i = 0, 1, 2, \cdots$, and maps $f_i \colon X_i \to \Omega X_{i+1}$. Our spaces are assumed to have base points and the maps are base point preserving. Also in this definition we take only loops of length one. The suspension below is the reduced suspension.

By the well-known natural isomorphism $\rho \colon \{X, \Omega Y\} \to \{sX, Y\}$, ($\{A, B\}$ = set of maps from A to B), the last is equivalent to being given maps $g_i \colon sX_i \to X_{i+1}$, where $g_i = \rho(f_i)$.

Example 1. Let X be any topological space and let $S(X)$ be the spectrum defined by the sequence of spaces $\{s^i X\}$, $i = 0, 1, 2, \cdots$ with maps $g_i \colon s(s^i X) \to s^{i+1}X$, $g_i =$ identity.

Definition 5.2. The homology groups $H_n(S)$ are defined by:

$$H_n(S) = \underset{\longrightarrow}{\mathrm{Lim}}\, H_{n+i}(X_i),\quad H_{n+i}(X_i) \to H_{n+i+1}(X_{i+1})$$

is the composite $H_{n+i}(X_i) \xrightarrow{\sigma} H_{n+i+1}(sX_i) \xrightarrow{g_i} H_{n+i+1}(X_{i+1})$, where σ is the suspension isomorphism.

The homotopy groups $\pi_n(S)$ are defined by:

$$\pi_n(S) = \underset{\longrightarrow}{\mathrm{Lim}}\, \pi_{n+i}(X_i),\quad \pi_{n+i}(X_i) \to \pi_{n+i+1}(X_{i+1})$$

is the composite $\pi_{n+i}(X_i) \xrightarrow{\sigma} \pi_{n+i+1}(sX_i) \xrightarrow{g_i} \pi_{n+i+1}(X_{i+1})$, where σ is the suspension homomorphism.

Example 2. If $S(X)$ is as in Example 1, then

$H_n(S(X)) = H_n(X)$, $n > 0$,
$H_0(S(X)) = H_0(X)$ (reduced homology group)
$\pi_n(S(X)) = n$-th stable homotopy group.

Definition 5.3. The *associated spectrum* $Q(S)$ *to the spectrum* S is defined by the sequence of spaces $\{s^i \Omega^i X_i\}$, $i = 0, 1, 2, \cdots$, and maps

$$s^{i+1}\Omega^i f_i \colon s(s^i \Omega^i X_i) \to s^{i+1}\Omega^{i+1} X_{i+1}.$$

Example 3. Let $S(X)$ be as in Example 1, then $Q(S(X))$ is defined by the sequence $\{s^i \Omega^i s^i X\}$ and maps $s^{i+1}\Omega^i f_i \colon s^{i+1}\Omega^i s^i X \to s^{i+1}\Omega^{i+1} s^{i+1} X$, where $f_i \colon s^i X \to \Omega s^{i+1} X$, $f_i = \rho^{-1}(g_i)$. The map f_i may be considered as an inclusion, and then $\Omega^i f^i \colon \Omega^i s^i X \to \Omega^{i+1} s^{i+1} X$ is an inclusion. Consequently, we may define $Q(X) = \mathrm{Lim}\, \Omega^i s^i X$. Then $Q(X)$ satsfies

$$H_n(Q(X)) = H_n(Q(S(X)))$$
$$\pi_n(Q(X)) = \pi_n(S(X))$$
$$\Omega Q(sX) = Q(X).$$

Also, one may consder the spectrum S' defined by $\{Q(s^i X)\}$ with maps $g'_i \colon Q(s^i X) \to \Omega Q(s^{i+1} X) = Q(s^i X)$, g'_i the identity. Then

$$H_n(Q(S')) = H_n(Q(X)) = H_n(Q(S(X))).$$

Example 4. Suppose that X is a space and we are given a map $f\colon X \to \Omega^k X$. Then we may form a spectrum S defined by the sequence: $X, \Omega^{k-1}X, \Omega^{k-2}X, \cdots, \Omega X, X, \Omega^{k-1}X, \cdots$, with every k-th map being $f\colon X \to \Omega(\Omega^{k-1}X)$, the other maps being identities.

If further, f is a homotopy equivalence (or at least induces isomorphisms of the homology groups), then

$$H_n(Q(S)) \cong H_n(X)$$

(*Remark.* This holds true of Bott's infinite Lie groups $U(\infty)$ and $SO(\infty)$.)

Let S *and* $Q(S)$ be as in Definitions 5.1 and 5.3. Then $H_n(Q(S)) = \underrightarrow{\mathrm{Lim}}\, H_{n+i}(s^i \Omega^i X_i) = \underrightarrow{\mathrm{Lim}}\, H_n(\Omega^i X)$, where in the last limit the maps are induced by $\Omega^i f_i \colon \Omega^i X_i \to \Omega^i(\Omega X_{i+1})$.

Since the inclusion of the subspace of loops of length one into the Moore loop space is natural and induces isomorphisms on homotopy and homology groups, we may substitute the Moore loop space in the last limit. Then $\Omega^i f_i$ is a map of H^{i-1}-spaces. This enables us to introduce homology operations $Q^{(p)}{}_j$, for arbitrarily large$j$, in $H_*(Q(S); Z_p)$.

Definition 5.4. For $\chi \in H_n(Q(S); Z_p)$, define $Q^{(p)}{}_j(\chi) = \mathrm{Lim}\, Q^{(p)}{}_j(\chi_i)$, where $\chi = \mathrm{Lim}\, \chi_i \in H_n(\Omega^i X_i)$, i sufficiently large so that $Q^{(p)}{}_j(\chi_i)$ defined.

This rich structure in $H_*(Q(S); Z_p)$ will enable us to compute $H_*(Q(S(X)); Z_p) = H_*(Q(X), Z_p)$ for any connected space X. We will show that it is the free graded algebra generated by all 'allowable words' in the homology operations, acting an a vector basis of $H^+{}_*(X; Z_p) \subset H_*(Q(X); Z_p)$. Explicitly,

Definition 5.5. A word mod p, p odd, is a formal product of Bocksteins $\partial^*{}_p$ and extended p-th powers $Q_{j(p-1)}$, $j > 0$. A word w acting on a class x of dim r is *allowable* if:

1. w is empty
2. $w = Q_{j(p-1)}w'$, w' allowable, and one of the following holds
   - (a) w' empty, j and r same parity
   - (b) w' begins with $Q_{i(p-1)}$, $j \leqq i$, j and i same parity
   - (c) w' begins with $\partial^*{}_pQ_{i(p-1)}$, $j < i$, j and i same parity
3. $w = \partial^*{}_pw'$, w' allowable, w' begins with $Q_{i(p-1)}$.

For any graded vector space V, we designate by $\mathfrak{F}^\infty(V)$ the free commutative, associative, graded algebra generated by all allowable words 'acting' on a basis of V^+.

We will show that there exists a natural inclusion

$$\phi_*: H_*(X;Z_p) \rightarrow H_*(Q(X);Z_p).$$

Since all operations are defined in this last homology algebra, ϕ_* extends to a homomorphism $\phi^\infty{}_*: \mathfrak{F}^\infty(H_*(X;Z_p)) \rightarrow H_*(Q(X;Z_p)$. Then it will remain to be shown that $\phi^\infty{}_*$ is an isomorphism.

Definition 5.6. A map $\phi: S \rightarrow S'$ of spectra, $S = \{X_i, f_i\}$, $S' = \{X'_i, f'_i\}$ is a sequence of maps $\phi_i: X_i \rightarrow X'_i$, such that $(\Omega, \phi_{i+1}) \circ f_i = f'_i \circ \phi_i$. In terms of the maps $g_i = \rho^{-1}f_i$, $g'_i = \rho^{-1}f'_i$, this condition is equivalent to $\phi_{i+1} \circ g_i = g'_i \circ (s\phi_i)$.

LEMMA 5.1. *The map* $\phi: S(X) \rightarrow Q(S(X))$ *defined by*

$$\phi_i: s^iX \rightarrow s^i\Omega^is^iX,\ \phi_i = s^i\rho^{-i}(s^i\lambda),\ \lambda: X \rightarrow X \textit{ the identity};$$

induces a monomorphism $\phi_*: H_*(S(X)) \rightarrow H_*(Q(S(X)))$.

Proof. The map

$$s^iX \xrightarrow{\phi_i} s^i\Omega^is^iX \xrightarrow{\psi_i} s^iX,\ \psi_i = \rho^i(\Omega^is^i\lambda),$$

is the identity. In fact,

$$\begin{aligned}\psi_i \circ \phi_i &= \rho^i(\Omega^is^i\lambda) \circ s^i\rho^{-i}(s^i\lambda) = \rho^i(\Omega^is^i\lambda \circ \rho^{-i}(s^i\lambda))\\ &= \rho^i\rho^{-i}(s^i\lambda \circ s^i\lambda) = s^i\lambda = \text{identity}.\end{aligned}$$

Consequently, $\phi^*{}_i: H_*(s^iX) \rightarrow H_*(s^i\Omega^isX)$ is a monomorphism for every i. It follows that ϕ_* is a monomorphism.

Remark. We had $H_*(X) \cong H_*(S(X))$ and $H_*(Q(X)) \cong H_*(Q(S(X)))$ and it is clear that the map $\phi_*: H_*(X) \rightarrow H_*(Q(X))$ is the monomorphism induced by the topological inclusion $X \rightarrow Q(X)$, defined by the sequence of inclusions $\rho^{-i}(s^i\lambda): X \rightarrow \Omega^is^iX$.

LEMMA 5.2. *Let X be connected. For p an odd prime, $H_*(\Omega^k s^i X; Z_p)$, $1 < k \leqq i$, is isomorphic in dimensions $< 3i - 2k$, to the free commutative associative graded algebra generated by:*

(1) *The allowable words in the extended p-th powers and the Bockstein, acting on a vector basis (x_l) of $H^+_*(s^{i-k}X; Z_p) \subset H_*(\Omega^k s^i X; Z_p)$; and*

(2) *The classes $\psi_{k-1}(x_l, x_m)$, $l < m$, where (x_l) is ordered by dimension (arbitrary order for elements of the same dimension); and $\psi_{k-1}(x_l, x_l)$, $\dim x_l$ same parity as k.*

Remark. If $x \in H_p(s^{i-k}X; Z_p)$ and $\dim Q_{j(p-1)}(x) < 3i - 2k$, then $j(p-1) < k$, and the operation is defined. In fact, since $r \geqq i - k + 1$, we have

$$3i - 2k > pr + j(p-1) \geqq pi - pk + j(p-1)$$

or

$$k > j(p-1) + (p-3)(i-k) + p > j(p-1).$$

Hence all allowable words in the operations within the dimension restriction are defined.

Proof. For $k = 2$, this is contained in the result of Browder [3]. (He obtains the complete algebra $H_*(\Omega^2 s^i X; Z_p)$, but his argument is quite long; we give a short argument in the proposition below for he weaker result required here.)

We now proceed by induction using the partial comparison theorem (4.1). That is, we can map the canonical spectral sequence into the spectral sequence for the acyclic fibre space $\Omega^{k+1}s^i X \subset P\Omega^k s^i X \rightarrow \Omega^k s^i X$ for terms of base degree $< 3i - 2k$; since for $k < i$, $H^+_*(s^{i-k}X; Z_p) \subset H_*(\Omega^k s^i X; Z_p)$ is transgressive, and hence the same is true for all operations on such classes. Thus, the base is transgressively generated in dimensions $< 3i - 2k$, and since $\Omega^k s^i X$ is an $H^{k-1}{}_p$-space, the hypothesis of Theorem (4.4) is fulfilled (see Theorem 4.7). It follows that $H_*(\Omega^{k+1}s^i X; Z_p)$ is isomorphic to the fibre of the canonical spectral sequence in dimensions $< 3i - 2k - 2 = 3i - 2(k+1)$.

Explicitly, $H_*(\Omega^{k+1}s^i X; Z_p)$ is generated by the transgressive images of the generators of $H_*(\Omega^k s^i X; Z_p)$, the transgressive images of the p-th powers of the generators of $H_*(\Omega^k s^i X; Z_p)$, and the Bocksteins of the latter. The words acting on a basis of $H^+_*(s^{i-k}X; Z_p)$ transgress into ($\pm$) the corresponding words acting on the corresponding basis of

$$H^+_*(s^{i-k-1}X; Z_p) \subset H_*(\Omega^{k+1}s^i X; Z_p);$$

i. e., $w(x_l)$ transgresses into $w'(\tau(x_l))$ where each $Q_{j(p-1)}$ in $w(x_l)$ is replaced in $w'(\tau(x_l))$ by $(\pm)Q_{(j+1)(p-1)}$, and Bocksteins transgress into Bocksteins. Also, $(w(x_l))^{p^f}$ transgresses into $Q^f{}_1w'(\tau x_l)$, and in addition we get the generators $\partial *_p Q^f{}_1 w'(\tau x_l)$. The words $w'(\tau(x_l))$, $Q^f{}_1(w'\tau x_l))$, $\partial *_p Q^f{}_1(w'(\tau x_l))$ give all allowable words (under the given restriction) in $H_*(\Omega^{k+1}s^iX;Z_p)$. Finally, each $\psi_{k-1}(x_l,x_m)$ transgresses into $\psi_k(\tau x_l,\tau x_m)$. (Notes that the p-th power of $\psi_{k-1}(x_l,x_m)$ is too large a dimension to enter the given range.)

This completes the induction and establishes the Lemma.

PROPOSITION 5.1. *Let X be connected. For p an odd prime, $H_*(\Omega^2s^iX;Z_p)$, $i \geqq 2$, is isomorphic in dimensions $\leqq 3i-4$, to the free commutative associative graded algebra generated by:*

(1) *The allowable words in the extended p-th power Q_{p-1} and the Bockstein, acting on a vector basis (x_l) of $H^+{}_*(s^{i-2}X;Z_p) \subset H_*(\Omega^2s^iX;Z_p)$; and*

(2) *The classes $\psi_1(x_l,x_m)$, $l<m$, (x_l) order by dimension; and $\psi_1(x_l,x_l)$, $\dim x_l$ even.*

Proof. $H_*(\Omega s^iX;Z_p)$ is well known (Cartan Seminar, 1954-55) to be the free associative (non-commutative) algebra generated by a basis (y_l) of $H^+{}_*(s^{i-2}X;Z_p) \subset H_*(\Omega s^iX;Z_p)$. Any such algebra may be written *additively* as a tensor product of polynomial algebras with new generators the 'basic products,' (b_r), $r=1,2,\cdots$, of the original generators (see Hilton or Browder [3] for details). If (y_l) is the original basis ordered by dimension, then the basic products are $y_1, y_2.\cdots$; $[y_l,y_m]$, all $l<m$; $[y_l,[y_m,y_n]]$, etc. Here $[y_l,y_m]=y_ly_m-(-1)^{\dim y_l \dim y_m}y_my_l=\psi_0(y_l,y_m)$. For any b_r of odd dimension we may further write the polynomial algebra $P(b_r) \cong E(b_r) \otimes P(b^2{}_r)$ (additive isomorphism), $E(b_r)$ the exterior algebra on b_r. The $H_*(\Omega s^iX;Z_p)$ will be additively isomorphic to the tensor product of exterior algebras generated by odd dim classes and polynomial algebras generated by even dimensional clases. Further, all these generators are transgressive. In fact, all y_l are transgressive and hence $[x_l,x_m]=\psi_0(x_l,x_m)$ is transgressive, etc. Also $\psi_0(b_r,b_r)=2b^2{}_r$, if b_r odd dimensional, and hence $b^2{}_r$ is transgressive.

Since Ωs^iX is an H-space, the differential in the spectral sequence of the fibration $\Omega^2s^iX \subset P\Omega s^iX \to \Omega s^iX$ is a derivation. Also

$$E^2 \cong H_*(\Omega s^iX;Z_p) \otimes H_*(\Omega^2s^iX;Z_p)$$

as algebras (in particular, any class in $H_*(\Omega e^iX;Z_p)$ commutes with any class in $H_*(\Omega^2s^iX;Z_p)$.

For dimensions $< 3i-2$, $H_*(\Omega s^i X; Z_p)$ is additively isomorphic to the tensor product of exterior algebras and polynomial algebras generated by the classes (y_l); $\psi_0(y_l, y_m)$, $l < m$; and $\psi_0(y_l, y_l)$, $\dim y_l$ odd. Further, since $\dim y_l \geqq i$, all p-th powers of the above classes are of dimension $\geqq 3i$. It follows that we may again map in our canonical spectral sequence (additively) for base degree $< 3i-2$. That is the images of the canonical sequence $C(b_S)$ for each basic product above are multiplied together in the given order in E. Since the differential commutes with the map on each $C^r(b_S)$ and the differential is a derivation in E^r, the map on $\otimes C(b_S)$ is a spectral sequence map (even though it is not an algebra map). Moreover the map is an algebra map on the fibre. The conclusion now follows by Theorem 4.1.

Remark. We have considered only p an odd prime. For $p=2$, the problem is much simpler because the operation $Q^{(2)}{}_1$ already exists in $H_*(\Omega^2 s^i X; Z_2)$, and hence all 2^f powers of transgressive classes in $H_*(\Omega s^i X; Z_2)$ again transgress. Similarly, the needed operations in the higher loop spaces exist. Consequently, one may apply the ordinary comparison theorem (and the canonical spectral sequence) to obtain the complete algebras $H_*(\Omega^k s^i X; Z_2)$ $k \leqq i$. Since this result is already given in Browder [3], we do not repeat it here.

Returning to Lemma 5.2, we see that for $k=i$, it states that $H_*(\Omega^i s^i X; Z_p)$ is isomorphic in dimensions $< i$ to the free commutative associative graded algebra generated by the allowable words acting on a vector basis of $H^+{}_*(X; Z_p) \subset H_*(\Omega^i s^i X; Z_p)$. (Note that $\dim \psi_{i-1}(x_l, x_m) > i$.) It follows that $\Omega^i p^{-1}(s^{i+1}\lambda): \Omega^i s^i X \to \Omega^{i+1} s^{i+1} X$ induces a homomorphism $\Omega^i p^{-1}(s^{i+1}\lambda)_*: H_r(\Omega^i s^i X; Z_p) \to H_r(\Omega^{i+1} s^{i+1} X; Z_p)$, $r < i$. Taking the direct limit we get:

THEOREM 5.1. *$H_*(Q(X); Z_p)$ is isomorphic to the free commutative associative graded algebra generated by the allowable words in the extended p-th powers and the Bockstein acting on a vector basis of $H^+{}_*(X; Z_p) \subset H_*(Q(X); Z_p)$.*

Remark. This result holds for $p=2$ as well as p an odd prime, as follows directly from the results of Browder discussed above.

COROLLARY. *$H_*(Q(sX); Z_p)$ is isomorphic as a Hopf algebra to the tensor product of monogenic Hopf algebras, these are exterior or polynomial algebras according as the generator is odd or even; the generators being the allowable words in the extended p-th powers and the Bockstein, acting on a vector basis of $H^+{}_*(sX; Z_p) \subset H_*(Q(sX); Z_p)$.*

Proof. This follows from the fact that all classes in $H_*(sX;Z_p)$ are primitive (in fact transgressive in $H_*(Q(sX);Z_p)$) and consequently the same is true of the words acting on these classes.

THEOREM 5.2. *Let S^n be a sphere of dimension n, n odd. Then $H_*(\Omega^k S^n, Z_p)$, $k < n$, is the tensor product as a Hopf algebra of monogenic algebras; exterior algebras for odd dim generators, polynomial algebras for even dim generators. Further there is one generator for each allowable word in the extended p-th powers $Q_{j(p-1)}$, $j < k$, and the Bockstein 'acting' on the generator of $H_{n-k}(\Omega^k S^n;Z_p)$.* (Note that we do not claim that these operations are actually defined in $\Omega^k S^n$.)

Proof. Consider the map $\Omega^k S^n \to \Omega^k Q(S^n) = Q(S^{n-k})$. We have $H_*(Q(S^{n-k});Z_p)$ is transgressively generated for $k < n-1$, and primitively generated for $k < n$. Further the map $\Omega S^n \to Q(S^{n-1})$ induces a monomorphism $H_*(\Omega S^n;Z_p) \to H_*(Q(S^{n-1});Z_p)$ since it takes generator into generator and both are free. Since $H_*(\Omega S^n;Z_p)$ is a polynomial algebra, the only indecomposable element is the generator, so that this map is also a monomorphism on the space I of indecomposable elements. By Theorem (4.9), it follows inductively, that

(1) $0 \to H_*(\Omega^k S^n;Z_p) \to H_*(Q(S^{n-k};Z_p)$ exact, $k < n$

(2) $0 \to I(H_*(\Omega^k S^n;Z_p) \to I(H_*(Q(S^{n-k};Z_p))$ exact, $k < n$

(3) $H_*(\Omega^k S^n;Z_p)$ is transgressively generated, $k < n-1$.

From (1) and (2) we conclude that $H_*(\Omega^k S^n;Z_p)$ is primitively generated and free, $k < n$. From (3) we see that the acyclic fibre space over $\Omega^k S^n$, $k < n-1$, has a canonical spectral sequence. Consequently, we can compute the number of generators in each successive loop space inductively. But it is easy to check that the number of generators must be the same as the number of allowable words and have the same dimensions.

COROLLARY. *$H_*(\Omega^k S^n;Z_p)$, $k < n$, n even, p an odd prime, is the tensor product as Hopf algebras of polynomial and exterior monogenic algebras. In fact, as Hopf algebras:*

$$H_*(\Omega^k S^n;Z_p) \cong H_*(\Omega^{k-1} S^{n-1};Z_p) \otimes H_*(\Omega^k S^{2n-1};Z_p).$$

Proof. The generator of $\pi_{2n-1}(S^n)$ defines a map $S^{2n-1} \to S^n$, from which we get a map of $S^{n-1} \times \Omega S^{2n-1} \to \Omega S^n$, by multiplying the inclusion $S^{n-1} \to \Omega S^n$ by the induced map $\Omega S^{2n-1} \to \Omega S^n$. This induces an isomorphism

$$\pi_i(S^{n-1} \times \Omega S^{2n-1}) \cong \pi_i(\Omega S^n)$$

in suitable C-theories (Serre [13]) and consequently the isomorphism of the Corollary.

THEOREM 5.3. *Let X be a homotopy commutative H-space, which is n-connected. If $H_*(\Omega^k X; Z_p)$ is primitively generated and all primitive elements transgress, $k \leqq n-1$; then $P^i u_q = 0$ for $i > (q-n)/2$, $u_q \in H^q(X; Z_p)$, P^i the Steenrod reduced power.*

Proof. The proof is by induction on q. If u_q is an element of lowest (non-zero) dimension in $H^* = H^*(X, Z_p)$. Then u_q is primitive, and likewise $P^i u_q$ is primitive. Since H_* is primitively generated, the p-th power of every element in H^* is zero, and a non-zero primitive element of H^* is indecomposable i. e.

$$0 \rightarrow P^* \rightarrow Q^*$$

is exact.

Since every primitive element in $H_*(\Omega^k X, Z_p)$ is the image of a primitive element in $H_*(\Omega^{k+1} X, Z_p)$, for $k < n-1$, $Q(H^*(\Omega^k X, Z_p))$ is mapped monomorphically into $Q(H^*(\Omega^{k+1} X, Z_p))$ by the suspension homomorphism. Also $Q(H^*(\Omega^{n-1} X, Z_p))$ is mapped monomorphically into $H^*(\Omega^n X, Z_p)$, since $H_*(\Omega^{n-1} X, Z_p)$ is transgressively generated. Hence if $s^n P^i u_q = 0$, then $P^i u_q = 0$. But $s^n P^i u_q = P^i s^n u_q = 0$ if $i > (q-n)/2$.

Now assume the theorem is true for all elements of dim less that q. Then if $u_q = v_r \cdot w_{q-r}$, $0 < r < q$, $P^i u_q = \sum_j P^j v_r \cdot P^{i-j} w_{q-r}$. Now by the induction assumption

$$\begin{aligned} &P^j v_r \neq 0 \text{ implies } j < (r-n)/2 \\ &P^{i-j} w_{q-r} \neq 0 \text{ implies } i-j < (q-r-n)/2 \\ \text{and} \quad &P^j v_r \cdot P^{i-j} w_{q-r} \neq 0 \text{ implies } i < (q-n)/2 + (-n)/2 < (q-n)/2. \end{aligned}$$

It follows that the assertion holds for any decomposable element. On the other hand, even if u_q is indecomposable we claim that $P^i u_q$ is primitive; and hence, zero by the same argument as in the first two paragraphs. To see that $P^i u_q$ is primitive, suppose $\phi^* u_q = u_q \otimes 1 + 1 \otimes u_q +$ terms of the form $v_r \otimes w_{q-r}$, $0 < r < q$ (ϕ^* the diagonal homomorphism in H^*). Again by the Cartan formula, and the argument above applied to $v_r \otimes w_{q-r}$, $P^i(v_r \otimes w_{q-r}) = 0$ if $i > (q-n)/2$. Hence $P^i u_q$ is primitive if $i > (q-n)/2$, and $P^i u_q = 0$.

The theorem now follows by induction.

COROLLARY 1. *In $H^*(Q(s^n X; Z_p)$, $P^i u_q = 0$ for $i > (q-n)/2$, X an arbitrary connected space.*

COROLLARY 2. *In* $H^*(\Omega^l S^{n+l+1}; Z_p)$; $P^i u_q = 0$ *for* $i > (q-n)/2$.

Proof. For $n+l+1$ odd, this follows from the fact that $H_*(\Omega^l S^{n+l+1}; Z_p)$ maps monomorphically onto $H_*(Q^{n+1}(S^{n+1}); Z_p)$ and hence the cohomology map coming back is onto. For even spheres it follows by Corollary to Theorem 5.2.

Remark. The above result may be sharpened slightly to give: In $H^*(\Omega^l S^{n+1}; Z_p)$, $P^i u_q = 0$ for $i > (q-n)/2$. The following is a brief outline of the argument:

All stable cohomology operations on $\gamma^*_n \in H^n(\Omega^l s^{n+l}; Z_p)$ are trivial since this is true in $H^*(S^{n+1}; Z_p)$. Hence the same is true of $\gamma^*_n \in H^n(Q(S^n); Z_p)$. On the other hand, all homology operations on $\gamma_n \in H_n(Q(S^n); Z_p)$ transgress to $H_{\ddagger}(Q(S^0); Z_p)$ (see Lemma 5.3 below); and hence as in the proof of Theorem 5.3, the result follows from the primitive cohomology classes and then for all classes by the Cartan formula. Hence as in Corollary 2, the result holds for $H^*(\Omega^l S^{n+l}; Z_p)$.

We wish to prove that certain cohomology operations are not zero in $H^*(\Omega^k S^n; Z_p)$. For this we need the:

LEMMA 5.3. *Let* X *be a k-th loop space and suppose* $x \in H_1(X; Z_p)$. *Then there exist a map* $\alpha_\pi : \dfrac{J^k\pi}{\pi} \to \Omega_0 X$, $\Omega_0 X$ *the connected component of the identity loop,* $\pi = Z_p$. *If* $\rho(\alpha_\pi) : s(J^k\pi/\pi) \to X$ *is the corresponding map, then* $(\rho\alpha_\pi)_*(s\bar{e}_{i+p-1}) = Q_i(x)$, $i+p-1 < k$, *where* $\bar{e}_{i+p-1} \in H_*(J^k\pi/\pi)$ *is* $t_{\pi *}(e_{i+p-1})$, $t_\pi : W^{(k-1)}/\pi \to J^k\pi/\pi$.

Proof. Since X is an $H^{k-1}{}_p$-space, we can define an $H^{k-1}{}_p$-structure on the space of paths PX by:

$$\bar{\theta}^{k-1}(w, \bar{x}_1, \cdots, \bar{x}_p)(t) = \theta^{k-1}(w, \bar{x}_1(t), \cdots, \bar{x}_p(t)),$$

$\bar{x}_i \in PX$, θ^{k-1} the defining map for X, $w \in J^{k-1}\pi$. This map restricted to ΩX gives an H^{k-1}-structure to ΩX which is the same as that given by the H^k-structure on ΩX defined previously (1.4). For any loop $w \in \Omega X$,

$$\alpha : J^k\pi \xrightarrow{\times w^p} J^k\pi \times w^p \xrightarrow{\theta^{k-1}} \Omega X \xrightarrow{w^{-p_0}} \Omega X$$

is equivariant, and the image of $J^k\pi$ will be in the identity component of ΩX. On the other hand, since $x \in H_1(X; Z_p)$ is spherical, it is represented by a loop w, and we define α as above for this w.

Now let X^* be the simply connected covering space of X, then the map $\theta^{k-1}{}_p : J^k\pi \times X^p \to X$ induces a unique lifted map $\theta^{k-1}{}_p : J^k\pi \times (X^*)^p \to X^*$, such that

$$\begin{tikzcd}
J^{k}{}_{\pi} \times (PX)^{p} \arrow[r, "\bar{\theta}"] \arrow[d, "1 \times \lambda_{*}"] & PX \arrow[r, "w^{-p_0}"] \arrow[d, "\lambda^{*}"] & PX \arrow[d, "\lambda^{*}"] \\
J^{k}{}_{\pi} \times (X^{*})^{p} \arrow[r, "\theta^{*}"] \arrow[d, "1 \times p^{p}"] & X^{*} \arrow[r, "w^{-p_0}"] \arrow[d, "p"] & X^{*} \arrow[dl, "p"] \\
J^{k}{}_{\pi} \times X^{p} \arrow[r, "\theta"] & X &
\end{tikzcd}$$ commutes.

The loop $w: [o, r] \to X$ defines a singular 1-cube in X, which we again denote by $x, x(t) = w(tr)$, $0 \leqq t \leqq 1$. Let $\bar{x}$ be the singular 1-cube in PX defined by $\bar{x}(t)(s) = w(ts)$, $0 \leqq t \leqq 1$, $0 < s \leqq r$. Then $\bar{x}$ projects on x by the end point map $\lambda: PX \to X$.

Consider $e_i \otimes \bar{x}^p$; this projects onto $e_i \otimes x^p$. We know from Section (III), that there exists a chain $\bar{c}$ in $W^{(n)} \otimes C(PX)^p$ of the form

$$e_i \otimes \bar{x}^p \pm \cdots \pm e_{i+p-1} \otimes (\partial \bar{x})^{p-1} \otimes \bar{x},$$

all terms except the first involving $\partial \bar{x}$, and such that $\partial \bar{c} = \pm e_{i+p-1} \otimes (\partial \bar{x})^p$. Now $\partial \bar{x} = w - \bar{c}$, so that the projection of the terms involving $\partial \bar{x}$ are terms involving $\partial x = 0$, and hence $(1 \otimes \lambda^p)_*(\bar{c}) = e_i \otimes x^p$. For then, $\bar{\theta}_*(\bar{c})$ is a chain in PX whose ∂ is in ΩX; hence $\partial \lambda^* \bar{\theta}_*(\bar{c})$ is degenerate in X^*; i. e. is a cycle, and $p\lambda^*(\bar{c}) = \lambda \bar{\theta}_*(\bar{c}) = \theta_*(\bar{c}) = Q_i(x)$.

On the other hand, we claim that $\bar{\theta}_*(e_{i+p-1} \otimes (\partial \bar{x})^p) = \bar{\theta}_*(e_{i+p-1} \otimes w^p)$ in ΩX. Once this is shown, we have $w_*^{-p} \bar{\theta}_*(e_{i+p-1} \otimes w^p) = \alpha_*(\bar{e}_{i+p-1})$, and our result will follow from the commutativity of the above diagram.

To show this last, note that the map $\theta^{k-1}{}_p$ as defined inductively for a k-th loop space satisfies the additional property.

$$\theta^{k-1}{}_p(w; x_1, \cdots, x_p) = \theta^{k-1}{}_p(w; e, \cdots, e, x_{i_1}, \cdots, x_{i_r}),$$

where $x_j \in X$, $j = 1, \cdots, p$, are such that $p - r$ of them are the base point e, the rest being $x_{i_1}, \cdots, x_{i_r}$ written in their given order; i. e., $i_1 < i_2 < \cdots < i_r$. This property then also holds of $\bar{\theta}^{k-1}{}_p: J^{k}{}_{\pi} \times (\Omega_0 X)^p \to \Omega_0 X$. For chains we claim we may then assume

$$\bar{\theta}(w \otimes \bar{x}_1 \otimes \cdots \otimes \bar{x}_p) = \bar{\theta}(w \otimes l \otimes \cdots \otimes l \otimes x_{i_1} \otimes \cdots \otimes \bar{x}_{i_r}),$$

where $\bar{x}_j \in C(\Omega_0 X)$, $j = 1, \cdots, p$, $p - r$ of them are the unit element l, the rest being $\bar{x}_{i_1}, \cdots, \bar{x}_{i_r}$ written in their given order. It follows immediately from this that all terms in the expansion of $\bar{\theta}_*(e_{i+p-1}\otimes(w-\bar{l})^p)$ involving the same number of $\bar{l}$'s are equal; and hence

$$\bar{\theta}_*(e_{i+p-1}\otimes(w-\bar{l})^p)=\bar{\theta}_*(e_{i+p-1}\otimes w^p)-\bar{\theta}_*(e_{i+p-1}\otimes\bar{l}^p), \bmod p.$$

But as remarked in (II.2), $\bar{\theta}_*(e_{i+p-1}\otimes\bar{l}^p)=0$, since $i+p-1>0$.

It remains to show that $\bar{\theta}$ has the desired property. To see this, for any space Y, take the explicit Σ_p-equivariant chain map $\phi\colon C(Y)^p\to C(Y^p)$ defined as follows: Let $g\colon C(Y)\to S(X)$ be the natural chain equivalence [5], $f\colon S(Y)^p\to S(Y^p)$ the explicit Eilenberg-Zilber map as given in [6], and $h\colon C(Y^p)\to S(X^p)$ be the natural chain equivalence inverse to g. Then since f is Σ_p-equivariant, the same is true of

$$\phi\colon C(Y)^p\xrightarrow{g^p}S(Y)^p\xrightarrow{f}S(Y^p)\xrightarrow{h}C(Y^p)$$

Then the map $t\otimes\phi\colon W^{(k-1)}\otimes C(Y)^p\to J^k\pi\otimes C(Y^p)$ satisfies the conditions of Theorem (2.1), and hence for $Y=\Omega_0X$ we may take $\bar{\theta}=\bar{\theta}\circ(t\otimes\phi)$. That $\bar{\theta}$ has the above property follows from the fact that ϕ is Σ_p-equivariant and is such that $\phi(e\otimes X^{p-1})\subset C(e\times X^{p-1})$, by the naturality of its component maps.

COROLLARY 1. *Let* $\gamma_{n-k}\in H_{n-k}(\Omega^kS^n;Z_p)$ *be a generator. There exists a map* $\rho^{n-k}(\alpha_\pi)\colon s^{n-k}(J^{n-1}\pi/\pi)\to\Omega^kS^n$ *such that*

$$\rho^{n-k}(\alpha_\pi)_*(s^{n-k}\bar{e}_{i+(n-k)(p-1)})=Q_i(\gamma_{n-k}),\ i<n-1-(n-k)(p-1).$$

COROLLARY 2. *Let* $Q(S^0)=\varinjlim\Omega^i{}_0S^i$. $H_*(Q(S^0);Z_p)$ *is the free associative commutative graded algebra with generators in* 1-1 *correspondence to all allowable words 'acting' on the unit element* $1\in H_*(Q(s^0);Z_p)$. (Note that we do not say $H_*(Q(s^0);Z_p)$ is generated by these words acting on 1, as all such operations are trivial.)

Proof. $Q(S^0)=\Omega_0Q(S^1)=\Omega(Q(S^1)^*)$, $Q(S^1)^*=$ simply connected covering space of $Q(S^1)$. $\pi_1(Q(S^1))=Z$, and hence by a well-known spectral sequence arguments we have the exact sequence of $H_*(Q(S^1)^*;Z_p)$ modules:

$$0\to H_*(Q(S^1)^*;Z_p)\to H_*(Q(S^1);Z_p)\to H_1(K(Z,1);Z_p)\otimes H_*(Q(S^1)^*;Z_p)\to 0.$$

It follows that $H_*(Q(S^1)^*;Z_p)$ is isomorphic to the subalgebra of $H_*(Q(S^1);Z_p)$ generated by all the non-empty allowable words acting on $H_*(S^1;Z_p)\subset H_*(Q(S^1);Z_p)$, but *not* the class γ_1 itself ($\gamma_1=$ generator of $H_*(S^1;Z_p)$).

Now the Lemma implies that all words of the form $Q_{j(p-1)}\gamma_1$ transgress, even though γ_1 no longer exists in $H_*(Q(S^1)^*;Z_p)$. Since all other allowable words are operations on these words, $H_*(Q(S^1)^*;Z_p)$ is transgressively generated. The Corollary now follows immediately.

As a consequence of Corollary 1 of Lemma 5.3, we show that there are a number of non-trivial ways in which the Steenrod Algebra can act on the cohomology of iterated loop spaces of spheres.

THEOREM 5.4. *In $H^*(\Omega^l S^{n+l};Z_p)$, if t is an even integer, $t \geqq n$ and $(t-n)(p-1) < l$, for which there is an integer j such that $(t/2(p-1),j) \not\equiv 0 \bmod p$ and $(t+2j)(p-1) < n+l-1$, then, letting $q = n+t(p-1)$, there is a class $u_q \in H^q(\Omega^l S^{n+l};Z_p)$ such that $P^j(u_q) \neq 0$. In particular, if $2j$ is an even multiple of $(p-1)$, $2jp < n+l-1$ and $0 \leqq 2j - n(p-1) < l$, then $P^j(u_{n+2j}) \neq 0$.*

Proof. By the cited Corollary there exists a map

$$\rho = \rho^{n-k}(\alpha_\pi) : s^n(J^{n+l-1}Z_p/Z_p) \to \Omega^l S^{n+1}$$

such that $\rho_*(s^n e_{i+n(p-1)}) = Q_i(\gamma_n)$, in mod p homology, provided $i < l$ and $i + n(p-1) < n+l-1$. Then if t is even and $t \geqq n$,

$$\rho_*(s^n e_{t(p-1)}) = Q_{(t-n)(p-1)}(\gamma_n) \neq 0.$$

Let $v \in H^2(J^{n+l-1}Z_p/Z_p;Z_p)$ denote the dual of (e_2). Then $v^{(t/2)(p-1)}$ is the dual of $e_{t(p-1)}$. An easy computation shows that

$$P^j(v^{t/2(p-1)}) = (t/2(p-1),j)v^{(t/2+j)(p-1)}, \ (t/2+j)(p-1) < n+l-1.$$

Thus, $P^j(s^n v^{t/2(p-1)}) = (t/2(p-1,j)\, s^n v^{(t/2+j)(p-1)}$. In $H^*(\Omega^l S^{n+1};Z_p)$, letting u_q be non-zero on $Q_{(t-n)(p-1)}(\gamma_n)$, we have $P^j(u_q) \neq 0$; provided

(1) $q = n + t(p-1)$,

(2) t is even and $t \geqq n$,

(3) $(t/2(p-1),j) \not\equiv 0 \bmod p$,

(4) $(t-n)(p-1) < l$, and

(5) $(t+2j)(p-1) < n+l-1$.

Letting $j = t/2(p-1)$, (3) is satisfied, (1) becomes $q = n+2j$, and (2), (4) and (5) are equivalent to

$$2jq < n+l-1, \text{ and } 0 \leqq 2j - n(p-1) < l.$$

Remark. Comparing Theorem 5.4 with Corollary 2 of Theorem 5.3 and the remark following it, we see that is is the best possible result.

REFERENCES.

[1] S. Araki and T. Kudo "Topology of H_n-spaces and H-squaring operations," *Memoirs of the Faculty of Science, Kyusyu University, Ser. A,* vol. 10 (1956), pp. 85-120.

[2] R. Bott, "The space of loops on a Lie group," *Michigan Mathematical Journal,* vol. 5 (1958), pp. 35-61.

[3] W. Browder, *Homology of loop spaces,* Thesis, Princeton, 1958.

[4] ———, "Suspensions in loop spaces," Abstract No. 553-79, *Notices of the American Mathematical Society,* vol. 5, no. 7 (1958), p. 820.

[5] S. Eilenberg and S. MacLane, "Acyclic models," *American Journal of Mathematics,* vol. 75 (1953), pp. 189-199.

[6] S. Eilenberg and J. Zilber, "On products of complexes," *ibid.*, pp. 200-204.

[7] Edward Halpern, *Twisted polynomial hyperalgebras,* Memoir No. 29, American Mathematical Society, 1958.

[8] I. M. James, "Reduced product spaces," *Annals of Mathematics,* vol. 62 (1955), pp. 170-197.

[9] J. Milnor, "Universal bundles II," *ibid.*, vol. 63 (1956), pp. 430-436.

[10] ———, "Geometric realization of a semi-simplicial complex," *ibid.*, vol. 65 (1957), pp. 357-362.

[11] J. Milnor and J. Moore, "The structure of Hopf algebrais," *to appear.*

[12] J. Moore, "The double suspension and p-primary components of homotopy groups of spheres," *Bol. de la Soc. Math. Mex.*, vol. 1 (1956), pp. 28-37.

[13] J. P. Serre, "Groupes d'homotopie et classes de groupes abélians," *Annals of Mathematics,* vol. 58 (1953), pp. 258-294.

[14] N. E. Steenrod, *The Topology of Fiber Bundles,* Princeton University Press, 1951.

[15] ———, "Homology groups of symmetric groups and reduced power operations," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 39 (1953), pp. 213-223.

[16] ———, "Cohomology operations derived from the symmetric group," *Commentarii Mathematici Helvetici,* vol. 31 (1957), pp. 195-218.

[17] R. G. Swan, "The homology of cyclic products," *Transactions of the American Mathematical Society,* vol. 95 (1960), pp. 27-68.

[18] G. W. Whitehead, "Homology theories and duality," *Proceedings of the National Academy of Sciences, S.A.*, vol. 46 (1960), pp. 554-556.

[19] E. C. Zeeman, "A proof of the comparison theorem, for spectral sequences," *Proceedings of the Cambridge Philosophical Society,* vol. 53 (1957), pp. 57-62.

ON GALOIS THEORY OF FIELDS WITH OPERATORS.*

By A. Białynicki-Birula.

0. Introduction. By a field with operators we mean a (commutative) field together with a family of automorphisms and derivations of the field containing the zero derivation d_0 and the identity automorphism ι (the last assumption is of purely technical character). As examples of fields with operators, we mention: (1) every differential field in the sense of Kolchin (p. 754 in [6]) together with d_0 and ι is a field with operators in our sense, (2) if G is a connected algebraic group defined over a field F then the field $F(G)$ of rational functions on G defined over F together with all automorphisms of $F(G)$ induced by left translations of G by F-rationals points $g \in G$ and with the zero derivation d_0 is a field with operators, (3) if G, F, $F(G)$ have the same meaning as in (2) then $F(G)$ together with the Lie algebra of all left invariant derivations of $F(G)$ and with the identity automorphism ι is a field with operators.

The main purpose of the present paper is to develop a Galois theory for fields with operators. The Galois theory of differential fields was originated and developed by Kolchin [4], [5], [6], and [7]. Theorems of Galois theory type are also known in the case where considered fields with operators are as in examples (2) ([1], [2]) and (3) ([2]). In all these cases we deal with a field with operators $\mathfrak{F}$ and an extension $\mathcal{E}$ of $\mathfrak{F}$ such that one can give a structure of an algebraic group to the group G of all automorphisms of $\mathcal{E}$ which leave the elements of $\mathfrak{F}$ fixed. Then there exists the usual one—one Galois correspondence between a family of algebraic subgroups of G and a family of subfields of $\mathcal{E}$ containing $\mathfrak{F}$. In best cases this is a correspondence between the family of all algebraic subgroups of G (defined over a certain field) and the family of all subfields of $\mathcal{E}$ (invariant with respect to the operators of $\mathcal{E}$) containing $\mathfrak{F}$. In this paper we shall also proceed along this way.

Section 1 contains definitions and notation used in this paper. In Section 2, we prove two useful lemmas on fields of constants of fields with operators. In Section 3, we give a definition of a strongly normal extension of a field with operators and we prove a theorem which links

---

* Received August 25, 1961.

the concept of strongly normal extensions with the theory of principal homogeneous spaces ([12]). It also shows that, in the case of strongly normal extension, the group of automorphisms over the ground field can be considered as an algebraic group. In Section 4, we prove a theorem on extension of isomorphisms and we show that, for strongly normal extensions, we have the usual one—one correspondence between the family of all connected algebraic subgroups of the group of automorphisms and the family of all subfields of a certain type of the extension (namely the family of all subfields $\mathcal{F}_1$ of the extension $\mathcal{E}$ such that $\mathcal{E}$ is a regular extension, in the usual sense, of $\mathcal{F}_1$ and which contains $\mathcal{F}$). Moreover, we prove that for differential fields (i. e., fields of characteristic 0 together with the identity automorphism and a family of derivations containing the zero derivation; we do not make any assumption on cardinality or commutativity of the family) we have the Galois correspondence between the family of all algebraic subgroups of the group of automorphisms and the family of all subfields of the extension containing the ground field. In Section 5, we give a definition of a Picard-Vessiot element and Picard-Vessiot extension. We show that Picard-Vessiot elements correspond to regular functions defined on a principal homogeneous space with respect to the group of automorphisms. Finally we give an application of this last result to the theory of ordinary differential fields.

This paper is greatly influenced by work of E. R. Kolchin ([6], [7], and with S. Lang [8]). This concerns not only the general line of this paper (suggested mainly by [8]) but also most of the given proofs of theorems and lemmas.

1. Notation and terminology. Let R be a commutative ring. Then we say that δ is an operator of R if δ is either an automorphism or a derivation of R. d_0 and ι will denote zero derivation and the identity automorphism of R, respectively. If R is an integral domain, then every operator δ of R can be uniquely extended to an operator of the field (R) of fractions of R. The extension will be also denoted by δ. Notice that the extensions of d_0 and ι are the zero derivation and the identity automorphisms of (R).

Let F be a subfield of two rings R_1 and R_2 and let δ_1 and δ_2 be operators of R_1 and R_2, respectively, such that $\delta_1(F) \subset F$, $\delta_2(F) \subset F$ and δ_1, δ_2 coincide on F. Then $\delta_1 \otimes_F \delta_2$ denotes the operator of $R_1 \otimes_F R_2$ defined by

$$(\delta_1 \otimes_F \delta_2)(\sum_i a_i \otimes b_i) = \begin{cases} \sum_i \delta_1(a_i) \otimes \delta_2(b_i) & \text{if } \delta_1, \delta_2 \text{ are automorphisms} \\ \sum_i (\delta_1(a_i) \otimes b_i + a_i \otimes \delta_2(b_i)) & \text{if } \delta_1, \delta_2 \text{ are derivations} \\ d_0 \text{ in the remaining cases.} \end{cases}$$

If P is a subring of R and α is a mapping of R into R then $\alpha|P$ denotes the mapping of P into R defined by $(\alpha|P)(a) = \alpha(a)$, for every $a \in P$. If C is a set of mappings of R into R then $C|P$ denotes the set of all mappings of the form $\alpha|P$, where $\alpha \in C$.

Let $\{\delta_\alpha\}_{\alpha \in A}$ be a set of operators of R containing the zero derivation d_0 and the identity automorphism ι of R. Then the pair $(R, \{\delta_\alpha\}_{\alpha \in A})$ is called a ring with operators.

In the case where R is an integral domain (a field) we shall also say that $(R, \{\delta_\alpha\}_{\alpha \in A})$ is an integral domain (a field) with operators. Rings (fields) with operators will be denoted by $\mathcal{R}, \mathcal{P}(\mathcal{E}, \mathcal{F}, \mathcal{F}_1, \mathcal{F}_1{}^0)$ and the rings (fields) corresponding to $\mathcal{R}, \mathcal{P}(\mathcal{E}, \mathcal{F}, \mathcal{F}_1, \mathcal{F}_1{}^0)$ will be denoted by $R, P(E, F, F_1, F_1{}^0)$, respectively.

The set $\{\delta_\alpha\}$ of operators of $\mathcal{R} = (R, \{\delta_\alpha\})$ is denoted by $O(\mathcal{R})$. $\mathbf{O}(\mathcal{R})$ denotes the set of all compositions of operators δ, where either $\delta \in O(\mathcal{R})$ or δ is an automorphism and $\delta^{-1} \in O(\mathcal{R})$.

We say that $\mathcal{R}$ is an extension of $\mathcal{P}$ or that $\mathcal{P}$ is a subring of $\mathcal{R}$ if R is an extension of P and $O(\mathcal{R})|P = O(\mathcal{P})$.

By a homomorphism of $\mathcal{R}_1 = (R_1, \{\delta_\alpha\}_{\alpha \in A})$ into $\mathcal{R}_2 = (R_2, \{\delta_\beta\}_{\beta \in B})$ we mean a pair (τ_1, τ_2) composed of a homomorphism τ_1 of R_1 into R_2 and a mapping τ_2 of A into B such that $\tau_1(\delta_\alpha(a)) = \delta_{\tau_2(\alpha)}(\tau_1(a))$, for every $a \in R_1$, $\alpha \in A$. In the case where $A = B$, a homomorphism τ of R_1 into R_2 is called a homomorphism of $\mathcal{R}_1$ into $\mathcal{R}_2$ if $\tau(\delta_\alpha(a)) = \delta_\alpha(\tau(a))$, for every $a \in R_1$. A homomorphism (τ_1, τ_2) is called an isomorphism if τ_1 is an isomorphism and τ_2 is one-one and onto. An isomorphism (τ_1, τ_2) is called an automorphism if $R_1 = R_2$ and τ_2 is the identity map of A. If $\mathcal{P}$ is a subring of $\mathcal{R}$ then $\mathcal{G}(\mathcal{R}/\mathcal{P})$ denotes the group of all automorphisms of $\mathcal{R}$ that leave all elements of $\mathcal{P}$ fixed.

We say that an element $c \in R$ is a constant of $\mathcal{R}$ if $\delta(c) = c\delta(1)$, for every $\delta \in O(\mathcal{R})$. If c is a constant of $\mathcal{R}$ then $\theta(ca) = c\theta(a)$ for every $\theta \in \mathbf{O}(\mathcal{R})$ and $a \in R$. Moreover, the set $\mathcal{R}_0(\mathcal{F}_0)$ of all constants of $\mathcal{R}(\mathcal{F})$ is a subring (a subfield) of $R(F)$ and it can be considered as a subring (a subfield) of $\mathcal{R}(\mathcal{F})$.

If $\mathcal{R} = (R, \{\delta_\alpha\}_{\alpha \in A})$ is an integral domain with operators then $\langle \mathcal{R} \rangle$ denotes $((R), \{\delta_\alpha\}_{\alpha \in A})$.

Let I be an ideal of R such that $\delta_\alpha(I) \subset I$, for every $\alpha \in A$. Then we say that I is an ideal of $\mathcal{R}$. $\delta_\alpha{}^I$ denotes the operator of R/I induced by δ_α and the ring with operators $(R/I, \{\delta_\alpha{}^I\}_{\alpha \in A})$ will be denoted by $\mathcal{R}/I$.

If $\mathcal{R}_1 = (R_1, \{\delta_\alpha\}_{\alpha \in A})$, $\mathcal{R}_2 = (R_2, \{\delta_\beta\}_{\beta \in B})$ are two extensions of $\mathcal{F}$ and $C \subset A \times B$ then $\mathcal{R}_1 \, {}_C\!\otimes_F \mathcal{R}_2$ denotes $(R_1 \otimes_F R_2, \{\delta_\alpha \otimes_F \delta_\beta\}_{(\alpha, \beta) \in C})$. We shall omit the subscript C in $\mathfrak{R}_1\, {}_C\!\otimes_F \mathfrak{R}_2$ if either $A = B$ and C is the diagonal of $A \times A$ or $\mathfrak{R}_2 = (R_2, \{d_0, \iota\})$ and $C = A \times B$.

Let E be a field and let F be a subfield of E. Then $E \otimes_F 1$ denotes the subring of $E \otimes_F E$ composed of all elements of the form $a \otimes 1$, where $a \in E$. $1 \otimes_F E$, $1 \otimes_F E \otimes_F E$, $E \otimes_F 1 \otimes_F 1$, etc., will have the analogous meaning. The isomorphisms of E onto $E \otimes 1$ and $1 \otimes E$ which send a into $a \otimes 1$, $1 \otimes a$, respectively will be called the canonical isomorphisms of E onto $E \otimes 1$ and $1 \otimes E$. For every $a \in F$, we shall identify a and $a \otimes 1 \in E \otimes_F E$. The analogous notation will be used for a field with operations $\mathcal{E}$ and a subfield $\mathcal{F}$ of $\mathcal{E}$.

If $\mathfrak{R}_1, \mathfrak{R}_2 (R_1, R_2)$ are two subrings of $\mathcal{E}(E)$ then $\langle \mathfrak{R}_1, \mathfrak{R}_2 \rangle$ $((R_1, R_2))$ denotes the smallest subfield of $\mathcal{E}(E)$ containing $\mathfrak{R}_1$ and $\mathfrak{R}_2$ (R_1 and R_2).

$\mathcal{E}$ is called a differential field if E is of characteristic 0 and δ is either a derivation or the identity automorphism, for every $\delta \in O(\mathcal{E})$. $\mathcal{E}$ is called a partial differential field if $\mathcal{E}$ is a differential field with a finite number of operators and $\delta_1\delta_2 = \delta_2\delta_1$ for any $\delta_1, \delta_2 \in O(\mathcal{E})$. A partial differential field is said to be an ordinary differential field if $0(\mathcal{E})$ contains at most one nontrivial derivation.

Let G be a connected algebraic group defined over a field F_0 and let V be a principal homogeneous space with respect to G defined over $F \supset F_0$. For any field $F_1 \supset F_0$ and $g \in G$, $G(F_1)$ denotes the group of all F_1-rational points of G, $F_1(g)$ denotes the field generated by g over F_1 and $F_1(G)$ denotes the field of rational functions on G defined over F_1. Similarly, for any $F_1 \supset F$ and $v \in V$, $F_1(v)$, $F_1(V)$ denote the field generated by v over F_1 and the field of rational functions on V defined over F_1, respectively. If $g \in G$, $F_1 \supset F$ then $\bar{g}_{F_1}$ denotes the automorphism of $F_1(g)(V)$ induced by the action of g on V. We shall often omit the subscript F_1 of $\bar{g}_{F_1}$. $\bar{G}(F_1)$ denotes the group of all automorphisms of $F_1(V)$ of the form $\bar{g}_{F_1}$, where $g \in G(F_1)$.

Examples 1. (a) Let V be a principal homogeneous space with respect to a connected algebraic group G, both defined over F. Then

$$(F(V), \{d_0\} \cup \{\bar{g}\}_{g \in G(F)})$$

is a field with operators.

(b) Let G be a connected algebraic group defined over F and let $\mathfrak{g}$ be the Lie algebra of G. Then $(F(G), \{\iota\} \cup \mathfrak{g})$ is a field with operators.

(c) Every differential field F in the sense of Kolchin (see, e.g., [6] p. 754), i.e., a field F of characteristic 0 together with a finite family of derivations $d_1, \cdots, d_m$ of F which commute in pairs, leads to a (uniquely determined) partial differential field (in the sense of this paper)

$$(F, \{\iota, d_0, d_1, \cdots, d_m\}).$$

In the sequel we shall identify differential fields in the sense of Kolchin and partial differential fields.

2. Lemmas on fields of constants.

LEMMA 1. *Let $\mathcal{E}$ be an extension of $\mathfrak{F}$. Then $\mathcal{E}_c$ and $\mathfrak{F}$ are linearly disjoint over $\mathfrak{F}_c$ and, for every subfield $\mathfrak{F}_1$ of $\mathcal{E}_c$, $\langle \mathfrak{F}, \mathfrak{F}_1 \rangle_c = \langle \mathfrak{F}_1, \mathfrak{F}_c \rangle$.*

Proof. Let us suppose that $\mathcal{E}_c$ and $\mathfrak{F}$ are not linearly disjoint over $\mathfrak{F}_c$ and let $a_1, a_2, \cdots, a_n$ be a minimal system of elements from $\mathcal{E}_c$ that are linearly independent over $\mathfrak{F}_c$ but are dependent over F. Then there exist $c_1 = 1, c_2, \cdots, c_n \in F$ such that

$$a_1 + c_2 a_2 + \cdots + c_n a_n = 0.$$

There exists i, $2 \leqq i \leqq n$, such that $c_i \in F - F_c$ and hence there exists $\delta \in O(\mathcal{E})$ such that $\delta(c_i) \neq c_i \cdot \delta(1)$. Then

$$\begin{aligned} &\delta(a_1) + \delta(c_2 a_2) + \cdots + \delta(c_n a_n) \\ &\qquad = \delta(1) a_1 + \delta(c_2) a_2 + \cdots + \delta(c_n) a_n = 0 \end{aligned}$$

and

$$\delta(1) a_1 + \delta(1) c_2 a_2 + \cdots + \delta(1) c_n a_n = 0.$$

Therefore

$$(\delta(c_2) - \delta(1) c_2) a_2 + \cdots + (\delta(c_n) - \delta(1) c_n) a_n = 0.$$

But $\delta(c_i) - \delta(1) c_i \neq 0$ and this contradicts the assumption of minimality of the system $a_1, \cdots, a_n$. Thus $\mathcal{E}_c$ and $\mathfrak{F}$ are linearly disjoint over $\mathfrak{F}_c$.

Now let $\mathfrak{F}_1$ be a subfield of $\mathcal{E}_c$. Then $\langle \mathfrak{F}, \mathfrak{F}_1 \rangle_c \supset \langle \mathfrak{F}_1, \mathfrak{F}_c \rangle$. Hence it suffices to prove that $\langle \mathfrak{F}, \mathfrak{F}_1 \rangle_c \subset \langle \mathfrak{F}_1, \mathfrak{F}_c \rangle$. Let $a \in \langle \mathfrak{F}, \mathfrak{F}_1 \rangle_c$ and let $\{f_\beta\}$ be an $\mathfrak{F}_c$-linear basis of F. Then a can be represented in the form $\sum_\beta a_\beta \cdot f_\beta / \sum_\beta b_\beta \cdot f_\beta$, where $a_\beta, b_\beta \in \langle \mathfrak{F}_1, \mathfrak{F}_c \rangle$. Therefore $\sum_\beta a_\beta \cdot f_\beta = \sum_\beta a b_\beta \cdot f_\beta$. But $a_\beta, a b_\beta \in \mathcal{E}_c$ and and it follows from the first part of the lemma that the elements $\{f_\beta\}$ are linearly independent over $\mathcal{E}_c$. Hence $a_\beta = a b_\beta$ and $a = a_\beta / b_\beta$ if $b_\beta \neq 0$. Thus $a \in \langle \mathfrak{F}_1, \mathfrak{F}_c \rangle$ since there exists β such that $b_\beta \neq 0$.

COROLLARY 1. *Let $\mathcal{E}$ be an extension of $\mathfrak{F}$ such that $\mathcal{E} = \langle \mathfrak{F}, \mathcal{E}_c \rangle$. If E is finitely generated over F then $\mathcal{E}_c$ is finitely generated over $\mathfrak{F}_c$.*

Proof. Let $a_1, \cdots, a_n$ be a finite set of generators of E over F. Then there exists a finite subset $b_1, \cdots, b_m$ of $\mathcal{E}_c$ such that every a_j, $1 \leqq j \leqq n$, can be expressed in the form $g(b_1, \cdots, b_m)$, where g is a rational function with coefficients from F. We shall prove that $b_1, \cdots, b_m$ generate $\mathcal{E}_c$ over $\mathfrak{F}_c$. Let $\{f_\beta\}$ be a linear basis of F over $\mathfrak{F}_c$ and let $a \in \mathcal{E}_c$. Then a can be represented in the form $\sum_\beta f_\beta \cdot g_\beta(b_1, \cdots, b_m) / \sum_\beta f_\beta \cdot h_\beta(b_1, \cdots, b_m)$, where g_β, h_β are polynomials with coefficients from $\mathfrak{F}_c$. Hence

$$\sum_\beta a f_\beta \cdot h_\beta(b_1, \cdots, b_m) = \sum_\beta f_\beta \cdot g_\beta(b_1, \cdots, b_m).$$

But $g_\beta(b_1, \cdots, b_m)$, $a h_\beta(b_1, \cdots, b_m) \in \mathcal{E}_c$, for every β, and it follows from Lemma 1 that $\{f_\beta\}$ are linearly independent over $\mathcal{E}_c$. Hence $a h_\beta(b_1, \cdots, b_m) = g_\beta(b_1, \cdots, b_m)$ and

$$a = g_\beta(b_1, \cdots, b_m) / h_\beta(b_1, \cdots, b_m)$$

unless $h_\beta(b_1, \cdots, b_m) = 0$. Thus $a \in \mathfrak{F}_c(b_1, \cdots, b_m)$ since there exists β such that $h_\beta(b_1, \cdots, b_m) \neq 0$.

LEMMA 2. *Let $a_1, a_2, \cdots, a_n \in \mathcal{E}$. Then the subspace of E^n spanned by all vectors of the form $(\theta a_1, \cdots, \theta a_n)$, where $\theta \in \mathbf{\Theta}(\mathcal{E})$, is different from E^n if and only if $a_1, a_2, \cdots, a_n$ are linearly dependent over $\mathcal{E}_c$.*

Proof. Suppose that $a_1, a_2, \cdots, a_n$ are linearly dependent over $\mathcal{E}_c$. Then there exist $c_1, c_2, \cdots, c_n \in \mathcal{E}_c$, not all equal to zero and such that $c_1 a_1 + c_2 a_2 + \cdots + c_n a_n = 0$. Hence $c_1 \theta(a_1) + \cdots + c_n \theta(a_n) = 0$, for every $\theta \in \mathbf{\Theta}(\mathcal{E})$. Therefore the ($E$-) subspace of E^n spanned by all vectors of the form $(\theta(a_1), \theta(a_2), \cdots, \theta(a_n))$ is different from E^n.

Conversely, suppose that the space generated by all vectors $(\theta(a_1), \cdots, \theta(a_n))$ is different from E^n. We may assume that E^{n-1} is spanned by

$$(\theta(a_1), \cdots, \theta(a_{n-1})), \theta \in \mathbf{\Theta}(\mathcal{E}).$$

There exist elements $c_1, c_2, \cdots, c_{n-1} \in E$ such that

$$c_1 \theta(a_1) + \cdots + c_{n-1} \theta(a_{n-1}) + \theta(a_n) = 0,$$

for every $\theta \in \mathbf{\Theta}(\mathcal{E})$. Hence

$$\delta(c_1 \theta(a_1)) + \cdots + \delta(c_{n-1} \theta(a_{n-1})) + (\delta\theta) a_n = 0,$$

for any $\delta \in O(\mathcal{E})$.

If δ is a derivation then this gives

$$\begin{aligned}0 = (\delta(c_1)\theta(a_1) + \cdots + \delta(c_{n-1})\theta(a_{n-1})) + (c_1(\delta\theta)(a_1) + \cdots \\ + c_{n-1}(\delta\theta)a_{n-1} + (\delta\theta)(a_n)) \\ - \delta(c_1)\theta(a_1) + \cdots + \delta(c_{n-1})\theta(a_{n-1}).\end{aligned}$$

Hence $\delta c_1 = \delta c_2 = \cdots = \delta c_{n-1} = 0$ since we have supposed that the space spanned by $(\theta(a_1), \cdots, \theta(a_{n-1}))$ is equal to E^{n-1}.

If δ is an automorphism then

$$\delta(c_1) \cdot (\delta\theta)(a_1) + \cdots + \delta(c_{n-1})(\delta\theta)(a_{n-1}) + (\delta\theta)(a_n) = 0.$$

Let us substitute $\theta = \delta^{-1}\theta_1$. Then we obtain

$$\delta(c_1)\theta_1(a_1) + \cdots + \delta(c_{n-1})\theta_1(a_{n-1}) + \theta_1(a_n) = 0,$$

for every $\theta_1 \in \mathfrak{O}(\mathcal{E})$.

On the other hand we have

$$c_1 \cdot \theta_1(a_1) + \cdots + c_{n-1} \cdot \theta_1(a_{n-1}) + \theta_1(a_n) = 0.$$

Hence $(c_1 - \delta(c_1))\theta_1(a_1) + \cdots + (c_{n-1} - \delta(c_{n-1}))\theta_1(a_{n-1}) = 0$ and so $c_j - \delta(c_j) = 0$ for $j = 1, \cdots, n-1$.

Thus $c_j \in \mathcal{E}_c$, for $j = 1, \cdots, n-1$, and $a_1, \cdots, a_{n-1}, a_n$ are linearly dependent over $\mathcal{E}_c$.

The following corollary follows easily from the above lemma.

COROLLARY 2. *Let $\mathfrak{R}$ be an integral domain with operators and let I be a prime ideal of $\mathfrak{R}$. Let τ be the canonical homomorphism of $\mathfrak{R}$ onto $\mathfrak{R}/I$ and let $\{a_1, \cdots, a_n\}$ be a finite subset of R. If the elements $a_1, \cdots, a_n$ are lineary dependent over $\langle\mathfrak{R}\rangle_c$ then $\tau(a_1), \cdots, \tau(a_n)$ are linearly dependent over $\langle\mathfrak{R}/I\rangle_c$.*

3. Strongly normal extensions.

Definition 1. Let $\mathcal{E}$ be an extension of $\mathcal{F}$. Then we say that $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}$ if

(i) E is a regular extension of F

(ii) E is finitely generated over F

(iii) $\mathcal{E}_c = \mathcal{F}_c$ and $\mathcal{F}_c$ is algebraically closed

(iv) $\langle\mathcal{E} \otimes_F \mathcal{E}\rangle = \langle\mathcal{E} \otimes_F 1, \langle\mathcal{E} \otimes_F \mathcal{E}\rangle_c\rangle$.

Examples 2. (a) If F is algebraically closed then

$$(F(V), \{d_0\} \cup \{\bar{g}\}_{g \in G(F)})$$

(defined in Examples 1 (a)) is a strongly normal extension of $(F, \{d_0, \iota\})$.

(b) If F is algebraically closed and of characteristic 0 then

$$(F(G), \{\iota\} \cup \mathfrak{g})$$

(defined in *Examples* 1(b)) is a strongly normal extension of $(F, \{d_0, \iota\})$.

(c) The following proposition gives another class of strongly normal extensions.

PROPOSITION 1. *Let a partial differential field $\mathcal{E}$ be an extension of $\mathfrak{F}$. Then $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ if and only if F is algebraically closed in E and $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ in the sense of* [6].

Proof. Let $\mathcal{E}$ be a strongly normal extension of $\mathfrak{F}$ in the sense of this paper. Then it follows from Definition 1 (i) that F is algebraically closed in E. Now, let $\mathfrak{F}^*$ be a universal extension of $\mathfrak{F}$ containing $\mathcal{E}$. Then there exists an isomorphism α of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle$ into $\mathfrak{F}^*$ such that $\alpha(a \otimes 1) = a$, for every $a \in E$. Let β be an isomorphism of $\mathcal{E}$ into $\mathfrak{F}^*$ defined by $\beta(a) = \alpha(1 \otimes a)$, for any $a \in E$. Then $\beta(a) = a$, for any $a \in F$, and β is isolated. But

$$\langle \alpha(\mathcal{E}), \mathcal{E} \rangle \approx \langle \mathcal{E} \otimes_F \mathcal{E} \rangle = \langle \mathcal{E}, \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c \rangle.$$

Hence β is a strong isomorphism. Therefore every $\mathfrak{F}$-isomorphism of $\mathcal{E}$ into $\mathfrak{F}^*$ is strong.

Thus $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ in the sense of [6].

Now suppose that F is algebraically closed in E and that $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ in the sense of [6]. Then conditions (i), (ii) and (iii) are satisfied evidently. Let $\mathfrak{F}^*$ and β have the same meaning as above. Then $\langle \mathcal{E}, \beta(\mathcal{E}) \rangle = \langle \mathcal{E}, \langle \mathcal{E}, \beta(\mathcal{E}) \rangle_c \rangle$ since β is strong. Hence

$$\langle \mathcal{E} \otimes_F \mathcal{E} \rangle = \langle \mathcal{E}, \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c \rangle$$

because $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle \approx \langle \mathcal{E}, \beta(\mathcal{E}) \rangle$. Thus condition (iv) is also satisfied. This completes the proof of Proposition 1.

THEOREM 1. *Let $\mathcal{E}$ be an extension of $\mathfrak{F}$ such that $\mathcal{E}_c = \mathfrak{F}_c$ and $\mathfrak{F}_c$ is algebraically closed. Then $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ if and only* *if there exists a connected algebraic group G defined over $\mathfrak{F}_o$ and a principal homogeneous space V for G defined over F, having the following properties:*

(a) *V is a model of E over F*

(b) $\mathfrak{G}(\mathcal{E}/\mathfrak{F}) \supset \bar{G}(\mathfrak{F}_o)$.

Conditions (a) *and* (b) *determine G and V uniquely up to an isomorphism. Moreover $\mathfrak{G}(\mathcal{E}/\mathfrak{F}) = \bar{G}(\mathfrak{F}_o)$ and G is a model of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$ over $\mathfrak{F}_o$ for every such G.*

Proof. Suppose that $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$. It follows from Corollary 1 that $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c$ is finitely generated over $\mathfrak{F}_o$. Let W be any model of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$ over $\mathfrak{F}_o$. This means that W is an algebraic variety defined over $\mathfrak{F}_o$ and that there exists a generic point w of W over $\mathfrak{F}_o$ such that $\mathfrak{F}_o(w) = \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$. Then $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$ is canonically isomorphic to $\mathfrak{F}_o(W)$ and we shall identify $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$ and $\mathfrak{F}_o(W)$ by means of this isomorphism.

$$\begin{aligned} \langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F \mathcal{E} \rangle &= \langle \mathcal{E} \otimes_F \langle \mathcal{E} \otimes_F 1, \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \rangle \rangle \\ &= \langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F 1, 1 \otimes_F \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \rangle \\ &= \langle \mathcal{E} \otimes 1 \otimes 1, \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes 1, 1 \otimes_F \langle \mathcal{E} \otimes \mathcal{E} \rangle_c \rangle. \end{aligned}$$

Thus it follows from Lemma 1 that

$$\langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F \mathcal{E} \rangle_o = \langle \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes 1, 1 \otimes \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \rangle.$$

Moreover $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes 1$ and $1 \otimes \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c$ are linearly disjoint over their intersection $\mathfrak{F}_o$. Let α, β, γ be the canonical isomorphisms of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o$ onto $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes 1$, $1 \otimes \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c$ and $\langle \mathcal{E} \otimes_F 1 \otimes_F \mathcal{E} \rangle_o$, respectively. Then

$$\langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F \mathcal{E} \rangle_c = \langle \alpha \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes_{\mathfrak{F}_o} \beta \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \rangle$$

and the mappings

$$\begin{aligned} &\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \xrightarrow[\gamma]{} \langle \mathcal{E} \otimes 1 \otimes \mathcal{E} \rangle_o \subset \langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F \mathcal{E} \rangle_o \\ &\quad = \langle \alpha \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes_{\mathfrak{F}_o} \beta \langle \mathcal{E} \otimes_F \mathcal{E} \rangle \rangle \xrightarrow[\alpha^{-1} \otimes \beta^{-1}]{} \langle \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \otimes_{\mathfrak{F}_o} \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_o \rangle \end{aligned}$$

determine a rational map $\boldsymbol{i}\colon W \times W \to W$ defined over $\mathfrak{F}_o$.

One can easily check that if w_1, w_2, w_3 are independent generic points of W over $\mathfrak{F}_o$ then

(G1) $\quad \mathfrak{F}_o(\boldsymbol{i}(w_1, w_2), w_1) = \mathfrak{F}_o(\boldsymbol{i}(w_1, w_2), w_2) = \mathfrak{F}_o(w_1, w_2)$

(G2) $\quad \boldsymbol{i}(\boldsymbol{i}(w_1, w_2), w_3) = \boldsymbol{i}(w_1, \boldsymbol{i}(w_2, w_3))$.

Therefore i is a normal law of composition on W and W with this law is a pre-group. The pre-group W is birationally equivalent to an algebraic group G_0 defined over $\mathfrak{F}_o$ (see Theorem p. 375 in [11]).

Now, let ϵ be the canonical isomorphism of E onto $1\otimes E \subset E\otimes_F E$ (defined by $\epsilon(a) = 1\otimes a$, for every $a\in E$) and let V be any model of E over F. Then the mappings $E \xrightarrow[\epsilon]{} 1\otimes E \subset E\otimes_F E = (E\otimes 1, \langle \mathcal{E}\otimes_F \mathcal{E}\rangle_o)$ determine a rational map $j\colon G_0\times V\to V$ defined over F, since $E\otimes 1$ and $(F\otimes 1, \langle\mathcal{E}\otimes_F\mathcal{E}\rangle_o)$ are linearly disjoint over F $(=F\otimes 1)$ and $E\otimes 1$, $(F\otimes 1, \langle\mathcal{E}\otimes_F\mathcal{E}\rangle_o)$ are canonically isomorphic to the fields of rational functions (defined over F) of V and G_0, respectively.

Again one can check that

(TG1′) If v, g are independent generic points of V, G_0 over F then $F(v,g) = F(v, j(g,v)) = F(g, j(g,v))$;

(TG2) If g_1, g_2, v are independent generic points of G_0, G_0 and V, respectively, taken over F, then $j(i(g_1,g_2),v) = j(g_1, j(g_2,v))$.

Finally G_0 and V have the same dimension. Therefore G_0 is simply pre-transitive on V. Hence by Proposition 3 in [12] V is birationally equivalent to a principal homogeneous space V_0 for G_0, defined over F.

Now we shall prove that $\bar{G}_0(\mathfrak{F}_o) = \mathfrak{G}(\mathcal{E}/\mathfrak{F})$. Lemmas 3 and 4 will provide a proof of this equality. A more general setting of the lemmas is dictated by their further applications.

Let $\mathcal{E}_1$ be an extension of $\mathcal{E}$ such that $\mathcal{E}_1 = \langle\mathcal{E}, (\mathcal{E}_1)_o\rangle$, i.e.,

$$\mathcal{E}_1 = \langle\mathcal{E}\otimes_{\mathcal{E}_o}(\mathcal{E}_1)_o\rangle.$$

Then E_1 is canonically isomorphic to $(F, (\mathcal{E}_1)_o)(V_0)$ (since $E = F(V_0)$). Hence every element $g\in G_0((\mathcal{E}_1)_o)$ induces an automorphism $\bar{g}_{(\mathcal{E}_1)_o}$ of E_1.

LEMMA 3. *$\bar{g}$ is an automorphism of $\mathcal{E}_1$.*

Proof. Let us identify $\mathcal{E}$ and $\mathcal{E}\otimes 1\subset\mathcal{E}\otimes_F\mathcal{E}$. It follows from the definition of j that there exists a point $g_0\in G_0(\langle\mathcal{E}\otimes_F\mathcal{E}\rangle_o)$ such that $\bar{g}_0(a) = \bar{g}_0(a\otimes 1) = 1\otimes a$, for every $a\in E$. Then $\bar{g}_0 \mid \mathcal{E}$ is an isomorphism into $\langle\mathcal{E}\otimes_F\mathcal{E}\rangle$ and g_0 is generic for G over E. Therefore, for every generic point $g\in G_0((\mathcal{E}_1)_o)$ of G_0 over E, $\bar{g}\mid E$ is an isomorphism of $\mathcal{E}$ into $\mathcal{E}_1$. Hence, for every such g, $\bar{g}$ is an automorphism of $\mathcal{E}_1$ since $\bar{g}\mid(\mathcal{E}_1)_o$ is the identity map and $\mathcal{E}_1 = \langle\mathcal{E}, (\mathcal{E}_1)_o\rangle = \langle\bar{g}(E), (\mathcal{E}_1)_o\rangle$ (notice also that E and $(\mathcal{E}_1)_o$ are linearly disjoint over $\mathcal{E}_o$). The case where g is arbitrary follows from the above and the fact that every element from G_0 can be represented as $g_1 \cdot g_2$, where g_1, g_2 are generic points of G_0 over E.

LEMMA 4. *Let* α *be an isomorphism of* $\mathcal{E}$ *into* $\mathcal{E}_1$ *such that* $\alpha(a) = a$, *for every* $a \in F$. *Then there exists* $g \in G_0((\mathcal{E}_1)_c)$ *such that* $\alpha = \bar{g} \mid \mathcal{E}$.

Proof. First, suppose that we have also an isomorphism β of $\mathcal{E}$ into $\mathcal{E}_1$ such that $\beta(a) = a$, for every $a \in F$, and $\alpha(E)$, $\beta(E)$ are linearly disjoint over F. We shall show that there exists an element $h \in G_0((\mathcal{E}_1)_c)$ such that $\bar{h}(a) = \beta\alpha^{-1}(a)$, for every $a \in \alpha(E)$. Indeed, let $\gamma : \langle \mathcal{E} \otimes_F \mathcal{E} \rangle \to \langle \alpha(\mathcal{E}), \beta(\mathcal{E}) \rangle$ be determined by $\gamma(\sum_i a_i \otimes b_i) = \sum_i \alpha(a_i) \cdot \beta(b_i)$, for every $a_i, b_i \in E$ and let $g_0 \in G_0$ be as in the proof of Lemma 3. Then γ is an isomorphism and $h = \gamma(g_0)$ satisfies the desired conditions.

Now consider $\langle \mathcal{E}_1 \otimes_F \mathcal{E} \rangle$. Then $\langle \mathcal{E}_1 \otimes_F \mathcal{E} \rangle = \langle \mathcal{E} \otimes 1, \langle \mathcal{E}_1 \otimes_F \mathcal{E} \rangle_c \rangle$ and it follows from the above result that there exist $g_1, g_2 \in G_0(\langle \mathcal{E}_1 \otimes_F \mathcal{E} \rangle_c)$ such that $\bar{g}_1(a \otimes 1) = 1 \otimes a$, $\bar{g}_2(\alpha(a) \otimes 1) = 1 \otimes a$, for every $a \in E$. Hence

$$\bar{g}_2^{-1} \cdot \bar{g}_1(a \otimes 1) = \bar{g}_2^{-1}(1 \otimes a) = \alpha(a) \otimes 1.$$

Moreover

$$F(g_2^{-1} g_1) \subset (E \otimes 1, \bar{g}_2^{-1} \cdot \bar{g}_1(E \otimes 1)) = (E \otimes 1, \alpha(E) \otimes 1) \subset E_1 \otimes 1.$$

Therefore $\mathfrak{F}_c(g_2^{-1} g_1) \subset (\mathcal{E}_1)_c$ and $g_2^{-1} \cdot g_1 \in G_0((\mathcal{E}_1)_c)$. Let $g = g_2^{-1} \cdot g_1$. Then $g \in G_0((\mathcal{E}_1)_c)$ and $\alpha = \bar{g} \mid \mathcal{E}$. This completes the proof of Lemma 4.

Taking $\mathcal{E}_1 = \mathcal{E}$ in Lemmas 3 and 4 we obtain immediately that $\mathfrak{G}(\mathcal{E}/\mathfrak{F}) = \bar{G}_0(\mathfrak{F}_c)$.

Now let us suppose that a conected algebraic group G and a principal homogeneous space V with respect to G satisfy conditions (a) and (b) of Theorem 1. We shall prove that $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$. In fact, (a) implies that E is a regular extension of F and E is finitely generated over F. Moreover, we have supposed that $\mathcal{E}_c = \mathfrak{F}_c$ and $\mathfrak{F}_c$ is algebraically closed. Therefore, it remains to show that

$$\langle \mathcal{E} \otimes_F \mathcal{E} \rangle = \langle \mathcal{E} \otimes 1, \langle \mathcal{E} \otimes \mathcal{E} \rangle_c \rangle.$$

Let g be any generic point of G over E and let $\mathcal{E}(g) = \langle \mathcal{E} \otimes_{\mathfrak{F}_c} \mathfrak{F}_c(g) \rangle$. Then $\mathcal{E}(g)_c = \mathfrak{F}_c(g)$ and $\mathcal{E}(g)$ is canonically isomorphic to $\mathfrak{F}(g)(V)$. Hence g induces a mapping $\bar{g}$ of $\mathcal{E}(g)$ into $\mathcal{E}(g)$. $\bar{g}$ is an automorphism of $\mathcal{E}(g)$. Indeed, consider the subset H of $G(\mathfrak{F}_c(g))$ composed of all elements h such that $\bar{h}$ is an automorphism of $\mathcal{E}(g)$. Then H is E-closed in $G(\mathfrak{F}_c(g))$, since $h \in H$ if and only if $h \in G(\mathfrak{F}_c(g))$ and $\delta(\bar{h}(f)) = \bar{h}(\delta(f))$, for every $f \in E$

and every operator $\delta \in O(\mathcal{E})$. On the other hand $G(\mathcal{F}_0) \subset H$ because we have assumed that $\bar{G}(F_0) \subset \mathfrak{G}(\mathcal{E}/\mathcal{F})$. Hence $H = G(\mathcal{F}_0(g))$, since $G(\mathcal{F}_0)$ is dense in G and so $g \in H$.

Now, E and $\bar{g}(E)$ are linearly disjoint over F (because g is generic for G over E). Therefore $\mathcal{E}(g) = \langle \mathcal{E}, \bar{g}(\mathcal{E}) \rangle$ is isomorphic to $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle$. But $\mathcal{E}(g) = \langle \mathcal{E}, \mathcal{F}_0(g) \rangle = \langle \mathcal{E}, \mathcal{E}(g)_0 \rangle$. Thus $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle = \langle \mathcal{E} \otimes 1, \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_0 \rangle$. This completes the proof that $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}$. We have proved also that $\mathcal{F}_c(g)$ is $\mathcal{F}_0$-isomorphic to $\langle \mathcal{E} \otimes \mathcal{E} \rangle_0$ and hence that G is a model of $\langle \mathcal{E} \otimes \mathcal{E} \rangle_0$ over $\mathcal{F}_0$.

Let us suppose again that G and V satisfy (a) and (b). Then we know that $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}$ and we can define the algebraic group G_0 and the principal homogeneous space V_0 with respect to G_0 as in the first part of the proof. We shall prove that G is isomorphic to G_0 and V to V_0 and this will give uniqueness of G and V and the equality $\bar{G}(\mathcal{F}_0) = \mathfrak{G}(\mathcal{E}/\mathcal{F})$, since we have proved that $\bar{G}_0(\mathcal{F}_c) = \mathfrak{G}(\mathcal{E}/\mathcal{F})$.

We shall identify E and $E \otimes 1 \subset E \otimes_F E$. Hence $(E \otimes_F E)$ is canonically isomorphic to $(F, \langle \mathcal{E} \otimes \mathcal{E} \rangle_c)(V)$. There exists a generic point g_0 of G over E such that $\mathcal{F}_0(g_0) = \langle \mathcal{E} \otimes \mathcal{E} \rangle_0$ and $\bar{g}_0(a) = \bar{g}_0(a \otimes 1) = 1 \otimes a$, for every $a \in E$. Indeed, let g be any generic point of G over E and let $\mathcal{E}(g) = \langle \mathcal{E} \otimes \mathcal{F}_c(g) \rangle$. Then, as we know, $\bar{g}$ is an automorphism of $\mathcal{E}(g)$. Let ϵ be the mapping of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle$ onto $\mathcal{E}(g)$ defined by $\epsilon(\sum_i a_i \otimes b_i) = \sum_i a_i \bar{g}(b_i)$. Then ϵ is an isomorphism and we can take $g_0 = \epsilon^{-1}(g)$.

Now, let g_1, g_2 be two independent generic points of G over E. Let $\mathcal{E}(g_1, g_2) = \langle \mathcal{E} \otimes_{\mathcal{F}_0} \mathcal{F}_0(g_1, g_2) \rangle$. Then $\bar{g}_1$, $\bar{g}_2$, $\bar{g}_2\bar{g}_1$ are automorphisms of $\mathcal{E}(g_1, g_2)$. Moreover E, $\bar{g}_1(E)$, $\bar{g}_2\bar{g}_1(E)$ are linearly disjoint over F. Therefore the mapping η of $\langle \mathcal{E} \otimes_F \mathcal{E} \otimes_F \mathcal{E} \rangle$ into $\mathcal{E}(g_1, g_2)$ defined by

$$\eta\Big(\sum_i a_i \otimes b_i \otimes c_i\Big) = \sum_i a_i \cdot \bar{g}_1(b_i) \cdot \bar{g}_2\bar{g}_1(c_i)$$

is an isomorphism. Let α_1, β_1, γ_1 be isomorphisms of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle$ into $\mathcal{E}(g_1, g_2)$ determined by

$$\alpha_1(a \otimes b) = a\bar{g}_1(b), \qquad \beta_1(a \otimes b) = \bar{g}_1(a) \cdot \bar{g}_2\bar{g}_1(b),$$
$$\gamma_1(a \otimes b) = a \cdot \bar{g}_2\bar{g}_1(b),$$

for every $a, b \in E$. Then $\alpha_1(g_0) = g_1$, $\beta_1(g_0) = g_2$, $\gamma_1(g_0) = g_2 g_1$ and

$$\eta\alpha = \alpha_1 | \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_c, \qquad \eta\beta = \beta_1 | \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_0, \qquad \eta\gamma = \gamma_1 | \langle \mathcal{E} \otimes_F \mathcal{E} \rangle_0,$$

where α, β, γ are the canonical mappings of $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_0$ onto $\langle \mathcal{E} \otimes_F \mathcal{E} \rangle_0 \otimes 1$,

$1\otimes\langle\mathcal{E}\otimes_F\mathcal{E}\rangle_c$ and $\langle\mathcal{E}\otimes 1\otimes\mathcal{E}\rangle_c$, respectively. Hence $g_2g_1=i(g_1,g_2)$, where i is the rational map of $G\times G$ into G defined at the beginning of the proof (G is a model of $\langle\mathcal{E}\otimes\mathcal{E}\rangle_c$ over $\mathfrak{F}_c$ and so we can apply the definition). Thus it follows from Theorem p. 375 in [11] that G is isomorphic to G_0 and we may assume that $G=G_0$. Let v be a generic point of V over F such that $F(v)=E$ and let g_0 be as above. Let $j'\colon G\times V\to V$ be the (external) law of composition of V with respect to G. Then $j'(g_0,v)=\bar{g}_0(v)$ and $\bar{g}_0(a\otimes 1)=1\otimes a$, for every $a\in E$. Hence j' coincides with j (defined for G_0 and V as at the beginning of the proof of Theorem 1) and so (by uniqueness of V_0, see Proposition 3 in [12]) V is isomorphic to V_0. This completes the proof of Theorem 1.

If $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ then the connected algebraic group G and the principal homogeneous space V with respect to G satisfying conditions (a) and (b) of Theorem 1 will be denoted by $G(\mathcal{E},\mathfrak{F})$ and $V(\mathcal{E},\mathfrak{F})$, respectively. We shall often write G and V instead of $G(\mathcal{E},\mathfrak{F})$ and $V(\mathcal{E},\mathfrak{F})$, when it does not lead to any ambiguity.

COROLLARY 3. *If $\mathcal{E}$ is a strongly normal extension of $\mathfrak{F}$ then for every $a\in E-F$ there exists $\alpha\in\mathfrak{G}(\mathcal{E}/\mathfrak{F})$ such that $\alpha(a)\neq a$.*

Proof. For every $a\in E-F$, there exists $g\in G(\mathfrak{F}_c)$ such that $\bar{g}(a)\neq a$ since $E=F(V)$. Hence by Theorem 1 we obtain the corollary.

Remark 1. One can easily see that in the case where $\mathcal{E}$ is a partial differential field $\mathfrak{G}(\mathcal{E},\mathfrak{F})$ is isomorphic to the algebraic group $\mathfrak{G}(\mathfrak{F})$ defined by Kolchin in [6] and [7].

4. Galois theory of strongly normal extensions.

LEMMA 5. *Let $\mathcal{E}$ be a strongly normal extension of $\mathfrak{F}$. Let I be a prime ideal of $\mathcal{E}\otimes_F\mathcal{E}$ and let τ be the canonical isomorphism of $\mathcal{E}\otimes_F\mathcal{E}$ onto $\mathcal{E}\otimes_F\mathcal{E}/I$. Then $\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle=\langle\tau(\mathcal{E}\otimes 1),\ \langle\tau(\mathcal{E}\otimes\mathcal{E})\rangle_c\rangle$.*

Proof. If $I=\mathcal{E}\otimes_F\mathcal{E}$ then the lemma is trivial. Suppose that $I\neq\mathcal{E}\otimes_F\mathcal{E}$. Then τ restricted to $\mathcal{E}\otimes 1$ is an isomorphism. Let $\{e_\beta\}$ be an $\mathfrak{F}_c$-linear basis of E. Then every element $a\in\mathcal{E}\otimes_F\mathcal{E}$ can be represented in the form $\sum\limits_\beta a_\beta(e_\beta\otimes 1)/\sum\limits_\beta b_\beta(e_\beta\otimes 1)$, where $a_\beta,b_\beta\in\langle\mathcal{E}\otimes_F\mathcal{E}\rangle_c$, since $\langle\mathcal{E}\otimes_F\mathcal{E}\rangle=\langle\mathcal{E}\otimes 1,\ \langle\mathcal{E}\otimes_F\mathcal{E}\rangle_c\rangle$. Hence $\sum\limits_\beta b_\beta\cdot a(e_\beta\otimes 1)=\sum\limits_\beta a_\beta(e_\beta\otimes 1)$ and so the elements $\{a(e_\beta\otimes 1),(e_\beta\otimes 1)\}$ are linearly dependent over $\langle\mathcal{E}\otimes_F\mathcal{E}\rangle_c$.

It follows from Corollary 2 that $\{\tau(a(e_\beta\otimes 1)),\tau(e_\beta\otimes 1)\}$ are linearly dependent over $\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_c$. Therefore there exist elements $a'_\beta, b'_\beta\in\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_o$ not all equal to zero such that $\sum_\beta b'_\beta\cdot\tau(a)\cdot\tau(e_\beta\otimes 1)=\sum_\beta a'_\beta\tau(e_\beta\otimes 1)$. But it follows from Lemma 1 that $\{\tau(e_\beta\otimes 1)\}$ are linearly independent over $\langle\tau(\mathcal{E}\otimes\mathcal{E})\rangle_o$ since $\{e_\beta\}$ are linearly independent over $\mathcal{E}_o$ and $\tau|\mathcal{E}\otimes 1$ is an isomorphism. Hence $\sum_\beta b'_\beta\cdot\tau(a)\cdot\tau(e_\beta\otimes 1)\neq 0$ unless $\tau(a)=0$. Therefore

$$\tau(a)=\sum_\beta a'_\beta\cdot\tau(e_\beta\otimes 1)/\sum_\beta b'_\beta\cdot\tau(e_\beta\otimes 1) \text{ unless } \tau(a)=0.$$

Thus $\tau(\mathcal{E}\otimes_F\mathcal{E})\subset\langle\tau(\mathcal{E}\otimes 1),\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_o\rangle$, and so

$$\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle=\langle\tau(\mathcal{E}\otimes 1),\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_o\rangle.$$

THEOREM 2. *Let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$ and let $\mathcal{F}_1$ be a subfield of $\mathcal{E}$ containing $\mathcal{F}$ such that E is a regular extension of F_1. Then every isomorphism α of $\mathcal{F}_1$ into $\mathcal{E}$ such that $\alpha(a)=a$, for every $a\in F$, can be extended to an automorphism of $\mathcal{E}$. Moreover $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}_1$.*

Proof. Let α be an isomorphism as in the theorem. Let I be the ideal of $E\otimes_F E$ generated by all elements of the form $a\cdot\alpha(b)\otimes c - a\otimes b\cdot c$, where $a, c\in E$, $b\in F_1$. Then I is a prime ideal (since E is a regular extension of F_1) and $\delta(I)\subset I$, for every $\delta\in O(\mathcal{E}\otimes_F\mathcal{E})$. Hence I is a prime ideal of $\mathcal{E}\otimes_F\mathcal{E}$. Let τ be the canonical homomorphism of $\mathcal{E}\otimes_F\mathcal{E}$ onto $\mathcal{E}\otimes_F\mathcal{E}/I$. Then τ restricted to $\mathcal{E}\otimes 1$ is an isomorphism and we shall identify $\mathcal{E}$ and $\tau(\mathcal{E}\otimes 1)$. It follows from Lemma 5 that $\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle=\langle\tau(\mathcal{E}\otimes 1),\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_o\rangle$ and from Lemma 4 that there exists $g\in G(\langle\tau(\mathcal{E}\otimes_F\mathcal{E})\rangle_o)$ such that $\bar g(\tau(a\otimes 1))=\tau(1\otimes a)$, for every $a\in E$. Then

$$\begin{aligned}\bar g(a\cdot b)=\bar g(\tau(ab\otimes 1))&=\tau(1\otimes ab)=\tau(\alpha(b)\otimes a)\\&=\tau(\alpha(b)\otimes 1)\cdot\tau(1\otimes a)=\alpha(b)\cdot\bar g(a),\end{aligned}$$

for every $a\in E$, $b\in F_1$. Now let $h\in G(\mathcal{F}_o)$ be a specialization of g over $\mathcal{F}_o$ then also $\bar h(ab)=\alpha(b)\bar h(a)$, for every $a\in E$, $b\in F_1$. Thus $\bar h(b)=\alpha(b)$, for every $b\in F_1$. Hence $\bar h_{\mathcal{F}_o}$ is an extension of α to an automorphism of $\mathcal{E}$.

Now, we shall show that $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}_1$. Conditions (i), (ii) and (iii) of Definition 1 are satisfied obviously. Moreover the equality $\langle\mathcal{E}\otimes_{F_1}\mathcal{E}\rangle=\langle\mathcal{E}\otimes 1,\langle\mathcal{E}\otimes_{F_1}\mathcal{E}\rangle_o\rangle$ follows from Lemma 5. This completes the proof of the theorem.

Let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. Then $\boldsymbol{F}_0(\mathcal{E},\mathcal{F})$ denotes the set of all subfields $\mathcal{F}_1$ of $\mathcal{E}$ containing $\mathcal{F}$ such that E is a regular extension of F_1. $\boldsymbol{F}(\mathcal{E}, \mathcal{F})$ denotes the set of all subfields of $\mathcal{E}$ containing $\mathcal{F}$.

$\boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$ $(\boldsymbol{G}(\mathcal{E}, \mathcal{F}))$ denotes the set of all connected algebraic subgroups (of all algebraic subgroups) of $G(\mathcal{F}_0)$ that are defined over $\mathcal{F}_0$.

Let, for any $\mathcal{F}_1 \in \boldsymbol{F}(\mathcal{E}, \mathcal{F})$, $\mathfrak{G}(\mathcal{F}_1)$ denotes the subgroup of all elements $g \in G(\mathcal{F}_0)$ such that $\bar{g}(a) = a$, for every $a \in F_1$. Then $\mathfrak{G}(\mathcal{F}_1) \in \boldsymbol{G}(\mathcal{E}, \mathcal{F})$.

Let, for any $G_1 \in \boldsymbol{G}(\mathcal{E}, \mathcal{F})$, $\mathfrak{F}(G_1)$ denotes the subfield of $\mathcal{E}$ composed of all elements $a \in E$ such that $\bar{g}(a) = a$, for every $g \in G_1$. Then $\mathfrak{F}(G_1) \in \boldsymbol{F}(\mathcal{E}, \mathcal{F})$.

THEOREM 3. *Let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. Then, for every $G_1 \in \boldsymbol{G}(\mathcal{E}, \mathcal{F})$, $\mathfrak{G}(\mathfrak{F}(G_1)) = G_1$, and for every $\mathcal{F}_1 \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$, $\mathfrak{F}(\mathfrak{G}(F_1)) = \mathcal{F}_1$. Moreover $\mathfrak{F}$ and $\mathfrak{G}$ establish the usual one-one Galois correspondence between $\boldsymbol{E}_0(\mathcal{E}, \mathcal{F})$ and $\boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$.*

Proof. For every $G_1 \in \boldsymbol{G}(\mathcal{E}, \mathcal{F})$, $\mathfrak{F}(G_1)$ is the field of rational functions on the space of G_1-orbits on V (see Theorem 2 in [9]). Hence $\mathfrak{F}(\mathfrak{G}(G_1)) = G_1$.

In order to prove that for every $\mathcal{F}_1 \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$, $\mathfrak{F}(\mathfrak{G}(\mathcal{F}_1)) = \mathcal{F}_1$ it suffices to notice that according to Theorem 2 $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}_1$ and so (by Corollary 3) for every $a \in E - F_1$ there exists $\alpha \in \boldsymbol{\mathfrak{G}}(\mathcal{E}/\mathcal{F}_1)$ such that $\alpha(a) \neq a$.

It remains to prove that, for every $\mathcal{F}_1 \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$, $\mathfrak{G}(\mathcal{F}_1) \in \boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$ and for any $G_1 \in \boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$, $\mathfrak{F}(G_1) \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$.

Let $\mathcal{F}_1 \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$. Then it follows from Theorems 2 and 1 that $\mathfrak{G}(\mathcal{F}_1)$ is connected and so $\mathfrak{G}(\mathcal{F}_1) \in \boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$.

Conversely, if $G_1 \in \boldsymbol{G}_0(\mathcal{E}, \mathcal{F})$ then $\mathcal{E}$ is a regular extension of $\mathfrak{F}(G_1)$ (see Theorem 2 in [9]) and so $\mathfrak{F}(G_1) \in \boldsymbol{F}_0(\mathcal{E}, \mathcal{F})$.

LEMMA 6. *Let $\mathcal{E}$ be a differential field and let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. Let $\mathcal{F}_1$ be a subfield of $\mathcal{E}$ containing $\mathcal{F}$. Then the algebraic closure $\mathcal{F}_1{}^0$ of $\mathcal{F}_1$ in $\mathcal{E}$ is a subfield of $\mathcal{E}$. Moreover every automorphism of $F_1{}^0$ over F_1 is an automorphism of $\mathcal{F}_1{}^0$ over $\mathcal{F}_1$ and $F_1{}^0$ is a normal extension of F_1.*

Proof. First, we shall prove that $\delta(F_1{}^0) \subset F_1{}^0$, for every $\delta \in O(\mathcal{E})$. If $\delta = \iota$ then this is obvious. Suppose that δ is a derivation. Then $\delta \mid F_1$ has a unique extension to a derivation δ^0 of $F_1{}^0$ into E (since E is of characteristic 0) and $\delta^0(F_1{}^0) \subset F_1{}^0$. But it follows from the uniqueness of δ^0 that $\delta^0 = \delta \mid F_1{}^0$ and that δ^0 commutes with every automorphism of $F_1{}^0$ over F_1.

Hence $\delta(F_1{}^0) \subset F_1{}^0$ and $\mathcal{F}_1{}^0$ is a subfield of $\mathcal{E}$. Moreover every automorphism of $F_1{}^0$ over F_1 is an automorphism of $\mathcal{F}_1{}^0$ over $\mathcal{F}_1$.

Let α be any isomorphism of $F_1{}^0$ into the algebraic closure $\bar{F}_1{}^0$ of $F_1{}^0$ which leaves the elements of F_1 fixed and let I be the ideal of $E \otimes_F E$ generated by all elements of the form $\sum_i a_i \otimes b_i$, where $a_i, b_i \in F_1{}^0$ and $\sum_i a_i \cdot \alpha(b_i) = 0$. Then I is prime (since $F_1{}^0$ is algebraically closed in E). Moreover, $\delta(I) \subset I$, for every $\delta \in O(\mathcal{E} \otimes_F \mathcal{E})$. In fact, $\delta = \delta_1 \otimes \delta_1$ for some $\delta_1 \in O(\mathcal{E})$. Every derivation $\delta' \in O(\mathcal{F}_1{}^0)$ can be extended in a unique way to a derivation of $\bar{F}_1{}^0$; the extended derivation commutes with α and it will be also denoted by δ'. ($\bar{F}_1{}^0$ with these derivations and the identity automorphism as operators will be denoted by $\bar{\mathcal{F}}_1{}^0$.) Therefore, if $a_i, b_i \in F_1{}^0$ and $\sum_i a_i \cdot \alpha(b_i) = 0$ then

$$0 = \delta'\Big(\sum_i a_i \cdot \alpha(b_i)\Big) = \sum_i \big(\delta'(a_i) \cdot \alpha(b_i) + a_i \cdot \alpha(\delta'(b_i))\big).$$

for every derivation $\delta' \in O(\bar{\mathcal{F}}_1{}^0)$. Hence

$$\delta\Big(\sum_i a_i \otimes b_i\Big) = \sum_i \big(\delta_1(a_i) \otimes b_i + a_i \otimes \delta_1(b_i)\big) \in I.$$

Thus I is a prime ideal of $\mathcal{E} \otimes_F \mathcal{E}$. Let τ be the canonical homomorphism of $\mathcal{E} \otimes_F \mathcal{E}$ onto $\mathcal{E} \otimes_F \mathcal{E}/I$. τ restricted either to $\mathcal{E} \otimes 1$ or to $1 \otimes \mathcal{E}$ is an isomorphism. We shall identify $\tau(\mathcal{E} \otimes 1)$ with $\mathcal{E}$ and $\tau(\mathcal{F}_1{}^0 \otimes_F \mathcal{F}_1{}^0)$ with $\langle \mathcal{F}_1{}^0, \alpha(\mathcal{F}_1{}^0)\rangle \subset \bar{\mathcal{F}}_1{}^0$ (according to the isomorphism which sends $\tau(a \otimes b)$ onto $a \cdot \alpha(b)$). It follows from Lemmas 4 and 5 that there exists an element $g \in G(\langle \tau(\mathcal{E} \otimes_F \mathcal{E})\rangle_o)$ such that $\bar{g}(a) = \bar{g}(\tau(a \otimes 1)) = \tau(1 \otimes a)$, for any $a \in E$. Then $\bar{g} \mid F_1{}^0 = \alpha$. Let $h \in G(\mathcal{F}_o)$ be any specialization of g over $\mathcal{F}_o$. Then $\bar{h}(\mathcal{E}) = \bar{h}(\tau(\mathcal{E} \otimes 1)) = \mathcal{E}$. On the other hand $\bar{h} \mid F_1{}^0 = \bar{g} \mid F_1{}^0$ since every element of $\bar{g}(F_1{}^0)$ is algebraic over $F_1{}^0$. Hence $\alpha(F_1{}^0) = \bar{g}(F_1{}^0) = \bar{h}(F_1{}^0) \subset E$. Thus $\alpha(F_1{}^0) = F_1{}^0$ and $F_1{}^0$ is a normal extension of F_1. This completes the proof of the lemma.

THEOREM 4. *Let $\mathcal{E}$ be a differential field and let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. Then $\mathfrak{F}$ and $\mathfrak{G}$ establish the usual one-one Galois correspondence between $\boldsymbol{F}(\mathcal{E}, \mathcal{F})$ and $\boldsymbol{G}(\mathcal{E}, \mathcal{F})$.*

Proof. If $\mathcal{F}_1 \in \boldsymbol{F}(\mathcal{E}, \mathcal{F})$, then $\mathfrak{F}\mathfrak{G}(\mathcal{F}_1) = \mathcal{F}_1$. In fact, let $F_1{}^0$ be the algebraic closure of F_1 in E. Then for every $a \in E - F_1{}^0$ there exists $\alpha \in \mathcal{G}(\mathcal{E}/\mathcal{F}_1)$ such that $\alpha(a) \neq a$, since (by Theorem 2) $\mathcal{E}$ is a strongly normal extension of $\mathcal{F}_1{}^0$. Let $a \in F_1{}^0 - F_1$. Then there exists an automorphism α_0 of $F_1{}^0$ over F_1 such that $\alpha_0(a) \neq a$ since (by Lemma 6) $F_1{}^0$ is a normal extension of F_1. By the last part of Lemma 6, α_0 is an automorphism of $\mathcal{F}_1{}^0$ over $\mathcal{F}_1$ and hence it follows from Theorem 2 that α_0 can be extended to an automorphism $\alpha \in \mathfrak{G}(\mathcal{E}/\mathcal{F}_1)$ such that $\alpha(a) \neq a$ and so $\mathfrak{F}(\mathfrak{G}(\mathcal{F}_1)) = \mathcal{F}_1$. The rest of Theorem 4 follows from Theorem 3.

5. Picard-Vessiot elements.

Definition 2. Let $\mathcal{E}$ be a field with operators and let F be a subfield of E. Then we say that an element $a \in E$ is a Picard-Vessiot element (PV-element) over F if the F-vector space spanned by all elements $\theta(a)$, where $\theta \in \boldsymbol{\Theta}(\mathcal{E})$, is finite dimensional. The dimension of the vector space is called the degree of a over F.

It follows from the above definition that

(i) if $a, b \in E$ are PV-elements over F then $a + b, a \cdot b$ are PV-elements over F

(ii) if $a \in E$ is a PV-element over F then, for every $\theta \in \boldsymbol{\Theta}(\mathcal{E})$, $\theta(a)$ is a PV-element over F.

Therefore the subset of E composed of all PV-elements over F is a subring of $\mathcal{E}$. This subring will be denoted by $\mathfrak{R}[\mathcal{E}, F]$.

LEMMA 7. *Let $\mathcal{E} = (E, \{\delta_\alpha\})$, $\mathcal{E}' = (E, \{\delta_\beta\})$ be two fields with operators such that $\delta_\alpha\delta_\beta = \delta_\beta\delta_\alpha$ for every α and β, and let $a \in E$. Then a is a PV-element of $\mathcal{E}$ over $\mathcal{E}'_0$ (of degree n) if and only if a is a PV-element of $\mathcal{E}'$ over $\mathcal{E}_0$ (of degree n).*

Proof. Let $a \in E$ be a PV-element of $\mathcal{E}$ over $\mathcal{E}'_0$ and let $\theta_1(a), \cdots, \theta_n(a)$, where $\theta_i \in \boldsymbol{\Theta}(\mathcal{E})$, for $i = 1, 2, \cdots, n$, be a basis of the $\mathcal{E}'_0$-vector space spanned by all elements of the form $\theta(a)$, where $\theta \in \boldsymbol{\Theta}(\mathcal{E})$. Let $\theta'_1, \cdots, \theta'_{n+1} \in \boldsymbol{\Theta}(\mathcal{E}')$ and consider the (E-) subspace of E^{n+1} generated by all vectors of the form $(\theta\theta'_1(a), \cdots, \theta\theta'_{n+1}(a))$, where $\theta \in \boldsymbol{\Theta}(\mathcal{E})$. This subspace is at most n-dimensional since it is spanned by vectors $(\theta_i\theta'_1(a), \cdots, \theta_i\theta'_{n+1}(a))$, for $i = 1, 2, \cdots, n$. In fact, $\theta(a)$ can be represented in the form $\sum_1^n c_i\theta_i(a)$ where $c_i \in \mathcal{E}'_0$ and so

$$\begin{aligned}(\theta\theta'_1(a), \cdots, \theta\theta'_{n+1}(a)) &= (\theta'_1\theta(a), \cdots, \theta'_{n+1}\theta(a)) \\ &= \sum_1^n (\theta'_1(c_i \cdot \theta_i(a)), \cdots, \theta'_{n+1}(c_i \cdot \theta_i(a))) \\ &= \sum_1^n c_i(\theta_i\theta'_1(a), \cdots, \theta_i\theta'_{n+1}(a)).\end{aligned}$$

Therefore it follows from Lemma 2 that $\theta'_1(a), \cdots, \theta'_{n+1}(a)$ are linearly dependent over $\mathcal{E}_0$. Thus the $\mathcal{E}_0$-vector space spanned by elements of the form $\theta'(a)$, where $\theta' \in \mathfrak{O}(\mathcal{E}')$, is at most n-dimensional and so a is a PV-element of $\mathcal{E}'$ over $\mathcal{E}_0$ of degree at most n. The rest of the lemma follows by symmetry.

THEOREM 5. *Let $\mathcal{E}$ be a strongly normal extension of $\mathfrak{F}$ and let $a \in E$. Then the following conditions are equivalent:*

(1) *a is a PV-element over F*

(2) *the $\mathfrak{F}_0$-vector space generated by $\mathfrak{G}(\mathcal{E}/\mathfrak{F})a$ is finite dimensional*

(3) *a is a regular function on $V(\mathcal{E}, \mathfrak{F})$.*

Proof. $(1) \Longleftrightarrow (2)$ Take $\mathcal{E}$ and $\mathcal{E}' = (E, \{d_0\} \cup \mathfrak{G}(\mathcal{E}/\mathfrak{F}))$, where d_0 is the zero derivation, and apply Lemma 7.

$(2) \Longrightarrow (3)$ Suppose that a is not regular and let V_1 be the algebraic subset of V composed of all points at which a is not defined. Then, for every point $v_0 \in G(\mathfrak{F}_0)V_1$, there exists a function from $\mathfrak{G}(\mathcal{E}/\mathfrak{F})a$ which is not defined at v_0. On the other hand $\bar{G}(\mathfrak{F}_0)V_1$ is dense in V. Therefore $\mathfrak{G}(\mathcal{E}/\mathfrak{F})a$ is not of finite dimension over F_0 (notice that, for every finite dimensional $\mathfrak{F}_0$-vector subspace of E $(=F(V))$, all functions of the space are defined on an open non-empty subset of V).

$(3) \Longrightarrow (2)$ Every regular function $f(g,v)$ on $G \times V$ defined over F can be represented in the form $\sum_i \bar{f}_i(g)\bar{e}_i(v)$, where $\bar{f}_i$, $\bar{e}_i$ are regular functions defined over F on G and V_1 respectively (see Exposé 2, p. 22, in [10]). If a is a regular function on V and g is a generic point over E then $\bar{g}(a)$ can be coinsidered as a regular function on $G \times V$. Therefore $\bar{g}(a) = \sum_i \bar{f}_i(g)\bar{e}_i(v)$, where $\bar{f}_i$, $\bar{e}_i$ are regular and defined over F. Moreover, every regular function f on G defined over F can be represented as $\sum_j c_j f_j$, where $c_j \in F$ and f_j are regular and defined over $\mathfrak{F}_0$. Hence $\bar{g}(a) = \sum_1^m f_i(g)e_i(v)$, where f_i are regular on G and defined over $\mathfrak{F}_0$ and e_i are regular on V and defined over F, for $i = 1, \cdots, m$. Thus, for every $h \in G(\mathfrak{F}_0)$, $\bar{h}(a) = \sum_1^m f_i(h)e_i$. But $\bar{G}(\mathfrak{F}_0) = \mathfrak{G}(\mathcal{E}/\mathfrak{F})$ and $f_i(h) \in \mathfrak{F}_0$ if $h \in G(\mathfrak{F}_0)$. Therefore $\mathfrak{G}(\mathcal{E}/\mathfrak{F}) \cdot a$ is contained in the $\mathfrak{F}_0$-vector space generated by e_i, $1 \leqq i \leqq m$. Thus the $\mathfrak{F}_0$-space spanned by $\mathfrak{G}(\mathcal{E}/\mathfrak{F})a$ is finite dimensional.

This completes the proof of Theorem 5.

Definition 3. Let $\mathcal{E}$ be a strongly normal extension of $\mathfrak{F}$. Then $\mathcal{E}$ is a PV-extension of $\mathfrak{F}$ if $\mathcal{E} = \langle \mathfrak{R}[\mathcal{E}, F] \rangle$.

The following Theorem follows easily from Theorem 5 ($(1) \Longrightarrow (3)$) and Definition 3.

THEOREM 6. *Let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. Then $\mathcal{E}$ is a PV-extension of $\mathcal{F}$ if and only if $G(\mathcal{E}, \mathcal{F})$ is affine.*

Examples 3. (a) Let G be connected algebraic group defined over an algebraically closed field F. For any $g \in G(F)$, let g_l denote the automorphism of $F(G)$ defined by $(g_l f)(h) = f(gh)$, for every $f \in F(G)$ and $h \in G$. Then $(F(G), \{d_0\} \cup \{g_l\}_{g \in G(F)})$ is a strongly normal extension of $(F, \{d_0, \iota\})$ and $\mathcal{R}[\mathcal{E}, F]$ coincides with the ring of representative functions of G (see p. 497 in [3]). The extension is a PV-extension if and only if G is affine.

(b) In the case where $\mathcal{E}$ is an ordinary differential field with derivation d then a is a PV-element over a subfield F of E if and only if a satisfies a linear homogeneous differential equation

$$d^n y + a_1 d^{n-1} y + \cdots + a_n y = 0, \text{ where } a_i \in F,$$

for $i = 1, \cdots, n$.

(c) The following Proposition gives a wide class of examples of PV-extensions.

PROPOSITION 2. *Let $\mathcal{E}$ be a partial differential field and let $\mathcal{E}$ be a strongly normal extension of $\mathcal{F}$. $\mathcal{E}$ is a PV-extension of $\mathcal{F}$ if and only if $\mathcal{E}$ is a Picard-Vessiot extension of $\mathcal{F}$ in the classical sense of* [5].

The above proposition follows from Theorem 6, Proposition 1, Remark 1 of this paper and Theorem 2, p. 891, in [7].

Let $\mathcal{E}$ be an ordinary differential field with derivation d and let $\mathcal{E}$ be a PV-extension of $\mathcal{F}$, say $\mathcal{E} = \mathcal{F}\langle e_1, \cdots, e_n\rangle$, where e_i, $1 \leqq i \leqq n$, are $\mathcal{F}_c$-independent solutions of a homogeneous linear equation.

$$d^n y + a_1 d^{n-1} y + \cdots + a_n y = 0,\ a_i \in F, \text{ for } i = 1, \cdots, n.$$

For $b_1, \cdots, b_m \in E$, let $\mathcal{F}\{b_1, \cdots, b_m\}$ denote the subring of $\mathcal{E}$ generated by $b_1, \cdots, b_m$ over F.

As an application of Theorem 5, we prove that

$$\mathcal{R}[\mathcal{E}, F] = \mathcal{F}\{e_1, \cdots, e_n, W(e_1, \cdots, e_n)^{-1}\},$$

where $W(e_1, \cdots, e_n)$ is the Wronskian of $e_1, \cdots, e_n$. We do not know any direct proof of this statement.

The inclusion $\mathfrak{F}\{e_1,\cdots,e_n\}\subset\mathfrak{R}[\mathcal{E},F]$ is trivial. Moreover,

$$d(W(e_1,\cdots,e_n))=-a_1W(e_1,\cdots,e_n)$$

and so $d(W(e_1,\cdots,e_n)^{-1})=a_1W(e_1,\cdots,e_n)^{-1}$. Thus $W(e_1,\cdots,e_n)^{-1}$ is also a PV-element over F and hence

$$\mathfrak{F}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}\subset\mathfrak{R}[\mathcal{E},F].$$

In order to prove the opposite inclusion, first suppose that F is algebraically closed. Then we may assume that $V(\mathcal{E},\mathfrak{F})=G(\mathcal{E},\mathfrak{F})\ (=G)$ and $F(G)=E$. For any $g\in G(\mathfrak{F}_o)$, $\bar{g}(e_i)=\sum_1^n c_{ij}(g)e_j$, where $c_{ij}(g)\in\mathfrak{F}_o$, for $i,j=1,\cdots.n$. Then c_{ij} are coordinate functions of an isomorphic rational matrix representation of G. Let e be the identity element of G and let $(b_{ij})_{i,j=1,\cdots,n}$ be a matrix of order n that in the i-th row and j-th column has b_{ij}, for any system $b_{ij}\in E$, $i,j=1,\cdots,n$. Then

$$((d^{i-1}e_j)(g))_{i,j=1,\cdots,n}\cdot((d^{i-1}e_j)(e))^{-1}{}_{i,j=1,\cdots,n}=(c_{ij}(g)),$$

for any $g\in G(\mathfrak{F}_o)$. Hence

$$(d^{i-1}e_j)_{i,j=1,\cdots,n}\cdot((d^{i-1}e_j)(e))^{-1}{}_{i,j=1,\cdots,n}=(c_{ij})_{i,j=1,\cdots,n}$$

and so

$$\mathfrak{F}\{\cdots,c_{ij},\cdots,\det(c_{ij})^{-1}\}\subset\mathfrak{F}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}.$$

But $\mathfrak{F}\{\cdots,c_{ij},\cdots,\det(c_{ij})^{-1}\}=\mathfrak{R}[\mathcal{E},F]$ (e.g. by Lemma 10.1 in [3]) and Theorem 5 $((1)\Longrightarrow(3))$. Hence

$$\mathfrak{R}[\mathcal{E},F]\subset\mathfrak{F}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}.$$

In the general case, let $\bar{\mathfrak{F}}$ be the algebraic closure of $\mathfrak{F}$ and let $\bar{\mathcal{E}}=\bar{\mathfrak{F}}\langle e_1,\cdots,e_n\rangle$. Then

$$\mathfrak{R}[\bar{\mathcal{E}},F]=\bar{\mathfrak{F}}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}.$$

$\bar{\mathfrak{F}}$ and $\mathcal{E}$ are linearly disjoint over $\mathfrak{F}$ since F is algebraically closed in E. Hence

$$\begin{aligned}\mathfrak{R}[\mathcal{E},F]=\mathfrak{R}[\bar{\mathcal{E}},F]\cap\mathcal{E}&=\bar{\mathfrak{F}}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}\cap\mathcal{E}\\&=\mathfrak{F}\{e_1,\cdots,e_n,W(e_1,\cdots,e_n)^{-1}\}\end{aligned}$$

and this completes the proof.

INSTITUTE OF MATHEMATICS,
POLISH ACADEMY OF SCIENCES, POLAND.

REFERENCES.

[1] E. Abe and T. Kanno, " Some remarks on algebraic groups," *Tohoku Mathematical Journal,* vol. 11 (1959), pp. 376-384.

[2] A. Białynicki-Birula, " On the field of rational functions of algebraic groups," *Pacific Journal of Mathematics,* vol. 11 (1961), no. 4, pp. 1205-1210.

[3] G. Hochschild and G. D. Mostow, " Representations and representative functions of Lie groups," *Annals of Mathematics,* vol. 66 (1957), pp. 495-542.

[4] E. R. Kolchin, "Algebraic matric groups and the Picard-Vessiot theory of homogeneous linear ordinary differential equations," *Annals of Mathematics,* vol. 49 (1948), pp. 1-42.

[5] ———, " Picard-Vessiot theory of partial differential fields," *Proceeding of the American Mathematical Society,* vol. 3 (1952), pp. 596-603.

[6] ———, " Galois theory of differential fields," *American Journal of Mathematics,* vol. 75 (1953), pp. 753-824.

[7] ———, " On the Galois theory of differential fields," *American Journal of Mathematics,* vol. 77 (1955), pp. 868-894.

[8] E. R. Kolchin and S. Lang, "Algebraic groups and the Galois theory of differential fields," *American Journal of Mathematics,* vol. 80 (1958), pp. 103-110.

[9] M. Rosenlicht, " Some basic theorems on algebraic groups," *American Journal of Mathematics,* vol. 78 (1956), pp. 401-443.

[10] Séminaire C. Chevalley (1956-1958), " Classification des groupes de Lie algèbrique," *Ecole Normal Supèrieure,* Paris, 1958.

[11] A. Weil, " On algebraic groups of transformations," *American Journal of Mathematics,* vol. 77 (1955), pp. 355-391.

[12] ———, " On algebraic groups and homogeneous spaces," *American Journal of Mathematics,* vol. 77 (1955), pp. 493-512.

SIMULTANEOUS QUADRATIC FORMS.*

By B. J. Birch, D. J. Lewis, and T. G. Murphy.

1. In this note we give an alternative proof of a theorem of Demyanov that two quadratic forms[1] in n variables over a $\mathfrak{p}$-adic field have a common $\mathfrak{p}$-adic zero provided n is at least nine (see [2]). Our proof is considerably simpler than Demyanov's—our main advantage arises from considering the pencil generated by the pair of forms, rather than considering the forms one at a time. There are two technical simplifications: the geometric crux, Lemma 2, is isolated from the rest of the argument; and the forms are reduced in a convenient way by minimising a function (see Lemma 3). We use the completeness of the $\mathfrak{p}$-adic field in Lemma 7 to show that it is sufficient to prove the theorem for 'general' pairs of forms.

2. For the moment, we work over a general field k. In this section, all forms will have coefficients in k, and when we refer to a zero we will mean a non-trivial zero with coordinates in k. We will say that a zero of a form, or of a pair of forms, is non-singular if the corresponding point is a non-singular point of the corresponding algebraic variety; that is to say, its tangent linear space is well-defined. In particular, $\mathfrak{X}$ is a non-singular zero of $f = f(x_1, x_2, \cdots, x_n)$ if the vector $\partial f(\mathfrak{X})/\partial x = \left(\dfrac{\partial f(\mathfrak{X})}{\partial x_1}, \cdots, \dfrac{\partial f(\mathfrak{X})}{\partial x_n}\right)$ is not the zero vector, and $\mathfrak{X}$ is a non-singular zero of the pair f, g if the vectors $f(\mathfrak{X})/\partial x$, $g(\mathfrak{X})/\partial x$ are not proportional.

If f is a form and T is a linear transformation, $f_T(x)$ is the form $f(Tx)$; denote the number of variables appearing explicitly in f by $\gamma(f)$; then the *order* $o(f)$ is defined by

$$o(f) = \min_T \gamma(f_T),$$

where the minimum is taken over all non-singular linear transformations T, defined over k. We say a form is degenerate if $o(f) < n$. We define the order of a pair of forms and a degenerate pair of forms, similarly.

---

* Received September 1, 1961.

[1] Form means homogeneous polynomial.

Of the lemmas which follow, Lemmas 1 and 3 are elementary and well-known. Lemma 2 is equally elementary but seems to be much less well-known; though Demyanov never states it explicitly, it is the crux of his proof of the theorem. It is the only step in our proof which may not easily be extended to cover the case of more than two forms.

LEMMA 1. *If a quadratic form has non-singular zeros in k, they cannot all lie in a proper linear subspace. A quadratic form can only have a singular zero if it is degenerate.*

Proof. Let $\mathfrak{X}$ be a non-singular point on a quadric and let π be a linear space through $\mathfrak{X}$. Take any point $\mathfrak{Y}$ with coordinates in k not in π and not in the tangent hyperplane at $\mathfrak{X}$ (this is possible even if k has only two elements). Then the line $\mathfrak{X}\mathfrak{Y}$ meets the quadric in another non-singular point not in π.

As for the last bit, a quadric with a singular point is a cone.

LEMMA 2. *If f, g is a non-degenerate pair of quadratic forms which have a common zero but no non-singular common zero, then there is a form of the pencil $\mu f - \lambda g$ which has only singular zeros.*

Proof. After a linear transformation we may assume that $e_1 = (1, 0, \cdots, 0)$ is a common zero of f and g. Since it is singular, the polar forms $f(e_1 | x)$ and $g(e_1 | x)$ are linearly dependent, so taking a suitable linear combination $h = \mu f - \lambda g$ we can make $h(e_1 | x)$ identically zero, thus x_1 does not occur in h. One of λ, μ is non-zero, say λ; then by the hypothesis that the pair f, g is not degenerate x_1 must appear in f, so after a linear transformation we may take

$$f = x_1 x_2 + f'(x_2, \cdots, x_n)$$
$$h = h(x_2, \cdots, x_n).$$

Since $\lambda \neq 0$, every common zero of f and g is a common zero of f and h and conversely.

If h had a non-singular zero in k, then by Lemma 1 it would have one with $x_2 \neq 0$. But then $(-f'/x_2, x_2, x_3, \cdots, x_n)$ would be a non-singular common zero of f and g, contrary to hypothesis.

LEMMA 3. *There is a polynomial $\mathfrak{d}(f, g)$ in the coefficients of f and g*

which is a combinant of the pencil, in the sense that if a, b, c, d are elements of k and T is a linear transformation then

$$\mathfrak{D}(af_T + bg_T, cf_T + dg_T) = (ad - bc)^{n(n-1)}(\det T)^{4(n-1)}\mathfrak{D}(f, g).$$

In fact, let F, G be the symmetric matrices associated with the forms f, g, and let $P(\lambda)$ be the characteristic polynomial,

$$P(\lambda) = \det |F - \lambda G|;$$

then we may take $\mathfrak{D}(f,g)$ as the discriminant of $P(\lambda)$,

$$\mathfrak{D}(f,g) = (\det G)^{2(n-1)} \prod_{i<j} (\lambda_i - \lambda_j)^2$$

where $\lambda_1, \cdots, \lambda_n$ are the roots of $P(\lambda) = 0$.

3. We now suppose that k is a $\mathfrak{p}$-adic field with ring of integers $\mathfrak{o}$, local prime π, prime ideal $\mathfrak{p} = \pi\mathfrak{o}$, and finite residue class field $k^* = \mathfrak{o}/\mathfrak{p}$. If a is in $\mathfrak{o}$, denote its canonical image in k^* by a^*. This homomorphism can be extended to a homomorphism of $\mathfrak{o}[x_1, \cdots, x_n]$ onto $k^*[x_1, \cdots, x_n]$; so for instance, if f is a form with integer coefficients then f^* denotes the residue class of $f \bmod \mathfrak{p}$; we say a vector $\mathfrak{X}$ over k is *primitive* if it has integer coordinates and $\mathfrak{X}^*$ is not the zero vector. Let $\nu(a)$ denote the greatest power of $\mathfrak{p}$ that divides a.

Before proceeding in the next section to the statement and proof of Demyanov's theorem, we quote or prove a number of easy lemmas.

LEMMA 4. *Every quadratic form f^* over the finite field k^* in at least three variables has a non-trivial zero in k^*; if $o(f^*) \geqq 3$ then f^* has a non-singular zero in k^*.*

LEMMA 5. *If f^*, g^* are a pair of quadratic forms over k^* in at least five variables then they have a non-trivial common zero in k^*.*

Both of these lemmas are particular cases of Chevalley's theorem on the solution of equations over finite fields (see [1], or up-to-date books on elementary number theory); to prove the last bit of Lemma 4 we express f^* in terms of the minimum number of variables and use Lemma 1.

LEMMA 6. *If f and g are quadratic forms over $\mathfrak{o}$ and if f^*, g^* have a non-singular common zero in k^* then f, g have a primitive common zero in k.*

Proof. This is essentially Hensel's Lemma. Let $\mathfrak{X}^{(1)}$ be the non-singular common zero in k^* of f^*, g^*; we will prove by induction on r that we can solve the simultaneous congruences

$$f(x) \equiv g(x) \equiv 0 \pmod{\mathfrak{p}^r} \qquad (E_r)$$
$$x \equiv \mathfrak{X}^{(1)} \pmod{\mathfrak{p}}.$$

Suppose that $\mathfrak{X}^{(r)}$ is a solution of (E_r); consider $\mathfrak{X}^{(r)} + \pi^r \mathfrak{Y}$, where $\mathfrak{Y}$ is integral and only matters mod $\mathfrak{p}$. Then

$$f(\mathfrak{X}^{(r)} + \pi^r \mathfrak{Y}) \equiv f(\mathfrak{X}^{(r)}) + \pi^r f(\mathfrak{X}^{(r)} | \mathfrak{Y}) \pmod{\mathfrak{p}^{r+1}}$$

where $f(x|y)$ is the polar form $\sum_{k=1}^{n} y_k \cdot \partial f / \partial x_k$. So to solve (E_{r+1}) it is enough to find $\mathfrak{Y}$ so that

$$\pi^{-r} f(\mathfrak{X}^{(r)}) + f(\mathfrak{X}^{(1)} | \mathfrak{Y}) \equiv \pi^{-r} g(\mathfrak{X}^{(r)}) + g(\mathfrak{X}^{(1)} | \mathfrak{Y}) \equiv 0 \pmod{\mathfrak{p}}.$$

Regarding these as linear equations for $\mathfrak{Y}$ over k^*, they are soluble since the vectors $\partial f^*(\mathfrak{X}^{(1)})/\partial x$, $\partial g^*(\mathfrak{X}^{(1)})/\partial x$ are linearly independent. For each r, let $\mathfrak{X}^{(r)}$ be a solution of (E_r) then any limit point $\mathfrak{X}$ of the set $\{\mathfrak{X}^{(r)}\}$ is a primitive common zero of f and g.

LEMMA 7. *Let $f^{(1)}, f^{(2)}, \cdots$ be a sequence of forms over k converging to the form f, and for each j let $\mathfrak{X}^{(j)}$ be a primitive zero of $f^{(j)}$. Then there is a subsequence such that the vectors $\mathfrak{X}^{(j_k)}$ converge to a primitive $\mathfrak{X}^{(0)}$ of f.*

This is clear: since the primitive vectors form a compact set, the $\mathfrak{X}^{(j)}$'s have at least one limit point, and this can only be a zero of f.

4. We are now in a position to prove

THEOREM 1. (Demyanov) *Let f, g be a pair of quadratic forms in $n \geqq 9$ variables over a $\mathfrak{p}$-adic field k; then they have a primitive common zero in k.*

Proof. We may assume without loss of generality that the coefficients of f, g are integers hence $\mathfrak{D}(f,g)$ is an integer. In the first three steps, we will suppose they are such that $\mathfrak{D}(f,g)$ defined in Lemma 3 is non-zero.

Step 1. Consider all pairs f', g' where f', g' are integer forms which can generate one of the pencils $\mu f_S - \lambda g_S$, S being any non-singular linear transformation over k; then f', g' are equivalent to f, g in the sense that there is a 1-1 correspondence between the common zeros of f', g' and the common zeros of f, g. There is a pair for which $\nu[\mathfrak{D}(f', g')]$ is minimal; without loss of generality we may suppose that f, g is such a pair from the start.

Step 2. We claim that (i) $o(f^*, g^*) \geqq 5$;

(ii) if h^* is in the pencil $\mu^* f^* - \lambda^* g^*$ then $o(h^*) \geqq 3$.

Indeed, if (i) were false, then there is a unimodular transformation U such that f_U^*, g_U^* involve at most the variables $x_1, \cdots, x_4$. Let R be the transformation that multiplies $x_1, \cdots, x_4$ by π and leaves the other $n-4$ variables alone; then $\pi^{-1} f_{UR}$, $\pi^{-1} g_{UR}$ are a pair of integer forms equivalent to f, g. But

$$[\mathfrak{D}(\pi^{-1} f_{UR}, \pi^{-1} g_{UR})] \\ = -2n(n-1) + 16(n-1) + \nu[\mathfrak{D}(f_U, g_U)] < \nu[\mathfrak{D}(f, g)],$$

contradicting the minimality of $\nu[\mathfrak{D}(f, g)]$.

Similarly if (ii) were false, suppose that $h^* = \mu^* f^* - \lambda^* g^*$ with $o(h^*) \leqq 2$ and say $\lambda^* \neq 0$. After a unimodular transformation U we may suppose that h_U^* involves at most x_1, x_2. This time, take R as the transformation that multiplies x_1, x_2 by π and leaves the others alone; then f_{UR}, $\pi^{-1} h_{UR}$ are a pair of integer forms equivalent to f, g and

$$\nu[\mathfrak{D}(f_{UR}, \pi^{-1} h_{UR})] \\ = -n(n-1) + 8(n-1) + \nu[\mathfrak{D}(f_U, \mu f_U - \lambda g_U)] < \nu[\mathfrak{D}(f, g)],$$

again a contradiction.

Step 3. Given (i) and (ii), it is easy to complete the proof. In fact, by (i) and Lemma 4, f^*, g^* have a common zero in k^*. If one of these zeros is non-singular, then by Lemma 6 we have a primitive common zero of f and g. Otherwise, by Lemma 2, there is a form h^* in the pencil of f^*, g^* which has only singular zeros; and, by Lemma 4, this contradicts (ii).

Step 4. Finally, we must deal with pairs f, g for which $\mathfrak{D}(f, g)$ vanishes. We can find a sequence $f^{(J)}$, $g^{(J)}$ of pairs with $\mathfrak{D}(f^{(J)}, g^{(J)})$ non-zero which converge $\mathfrak{p}$-adically to f, g as $J \to \infty$. If we wish to be unnecessarily explicit, we can take $F^{(J)} = F + \pi^J I$ and $G^{(J)} = G + \pi^J D$ where I is the unit matrix and D a diagonal matrix with all its diagonal elements different; for

$$\mathfrak{d}(f + \alpha \sum x_i^2, g + \alpha \sum d_i x_i^2)$$

is a polynomial in α which is not identically zero (the coefficient of the highest power of α being $\mathfrak{d}(\sum x_i^2, \sum d_i x_i^2) \neq 0$) and therefore does not vanish for sufficiently small values of α, e. g. $\alpha = \pi^J$, J large. By the general case of the theorem, each pair $f^{(J)}$, $g^{(J)}$ have a primitive zero, $\mathfrak{X}^{(J)}$ say. Applying Lemma 7, the $\mathfrak{X}^{(J)}$ have a convergent subsequence tending to a primitive vector $\mathfrak{X}^{(0)}$ which is a common zero for f and g.

CHURCHILL COLLEGE, CAMBRIDGE,
UNIVERSITY OF MICHIGAN,
ST. CATHERINE'S COLLEGE, CAMBRIDGE.

REFERENCES.

[1] C. Chevalley, "Démonstration d'une hypothèse de M. Artin," *Abh. Math. Seminar Hamburg*, vol. 11 (1935), pp. 73-75.

[2] V. B. Demyanov, "Pairs of quadratic forms over a complete field with discrete norm with a finite residue class field," *Izv Akad. Nauk SSSR*, vol. 20 (1956), pp. 307-324 (Russian).

ON THE PURITY OF THE BRANCH LOCUS.*[1]

By Maurice Auslander.

Let R be a regular local ring with field of quotients K, let L be a finite separable field extension of K and S the localization of the integral closure of R in L at some maximal ideal. If $\mathfrak{m}$ and $\mathfrak{M}$ are the maximal ideals of R and S respectively, we have that $\mathfrak{M} \supset \mathfrak{m}$ and $S/\mathfrak{M}$ is a finite dimensional field extension of $R/\mathfrak{m}$. The purity of the branch locus states that if every minimal prime ideal in S is unramified over R, then S is unramified over R (see Section 1 for the definition of unramified). In the geometric context this theorem is due to O. Zariski [11]. M. Nagata in [9] gave a proof of the theorem as stated above based on Chow's local Bertini theorem [7]. It is our purpose to give another proof of this theorem based upon the following module theoretic fact established in this note: If R is a regular local ring and M is a reflexive module such that $\mathrm{Hom}_R(M,M)$ is isomorphic to a direct sum of copies of M, then M is a free R-module.

1. Reduction to theorem on modules. In this section we assume that all local rings are integrally closed, noetherian domains. Let R be a local ring with field of quotients K. We shall say that a local ring $S \supset R$ is an R-algebra if: a) The field of quotients L of S is a finite dimensional separable extension of K contained in an algebraic closure of K; b) S is the localization at some maximal ideal of the integral closure of R in L. It follows immediately that the maximal ideal $\mathfrak{M}$ of S contains the maximal ideal $\mathfrak{m}$ of R and that $S/\mathfrak{M}$ is a finite dimensional field extension of $R/\mathfrak{m}$. We shall say that S *is an unramified R-algebra* if $\mathfrak{m}S = M$ and $S/\mathfrak{M}$ is a separable field extension of $R/\mathfrak{m}$.

Let $\mathfrak{P}$ be a prime ideal in S and let $\mathfrak{p} = \mathfrak{P} \cap R$. Then $\mathfrak{p}$ is a prime ideal in R and $S_{\mathfrak{P}}$ is an $R_{\mathfrak{p}}$-algebra. It is well known (see [3] for instance) that if S is an unramified R-algebra, then $S_{\mathfrak{P}}$ is an unramified $R_{\mathfrak{p}}$-algebra for all prime ideals $\mathfrak{P}$ in S. The problem of the "purity of the branch locus" is to investigate when a local ring R has the property that given any R-algebra S whose minimal prime ideals are unramified (i.e., given any minimal prime

---

* Received October 3, 1961.

[1] This work was done while the author was supported by a Senior Post-doctoral N. S. F. Fellowship.

ideal $\mathfrak{P}$ in S, the $R_{\mathfrak{p}}$-algebra $S_{\mathfrak{P}}$ is unramified), then S is unramified. We shall say that purity holds for R if R has this property for R-algebras. It is our purpose in this section to show that if a certain module theoretic statement holds for certain unramified R-algebras, then purity holds for R. (See Proposition 1.2).

Let S be an R-algebra with field of quotients L and let M be the galois closure of L in an algebraic closure of K chosen once and for all. Let T be an S-algebra whose field of quotients is M. Let G be the galois group of M over K and let G_I be the inertial group of T, i.e., the subgroup of G consisting of those elements $g \in G$ such that $g(T) = T$. If M_I is the fixed field of G_I and $U = T \cap M_I$, then it is well known (see [1] for instance), that

a) U is an unramified R-algebra with the same residue field as R

b) T is a finite U-algebra (i.e., T is a finitely generated U-module)

c) the composite SU or S and U in T is an unramified S-algebra and a finitely generated U-module.

Now suppose every minimal prime ideal in S is unramified over R. Then since SU is an unramified S-algebra, every minimal prime ideal of SU is unramified over S. Therefore every minimal prime ideal of SU is unramified over R (by transitivity of unramified extensions). Thus every minimal prime ideal of SU is unramified over U. Consequently if the R-algebra S has the property that every minimal prime ideal is unramified over R, then the finite U-algebra SU has the same property with respect to U. Suppose this property of SU over U implies that SU is unramified over U. Then SU is unramified over R (since U is unramified over R). Since $SU \supset S \supset R$ and SU is an unramified R-algebra while S is an R-algebra (remember, integrally closed), it follows that S is an unramified R-algebra (see [6] and [8]). Consequently purity holds for R if purity holds for U with respect to finite U-algebras.

Now M is a galois extension of K_I, the field of quotients of U, with group G_I. Since M is the least galois extension of L, it is immediate that M is the least galois extension of LK_I (the composite of L and K_I) which is the field of quotients of SU. Therefore T is the composite of the automorphic images of SU. Therefore if every minimal prime ideal in SU is unramified over U, every minimal prime ideal in T is also unramified over U, since the composite of unramified extensions is unramified. Suppose this property of T with respect to U forced T to be an unramified U-algebra. Then since $T \supset SU \supset U$ and SU is a U-algebra, we have that SU is an unramified U-algebra. Thus if purity holds with respect to finite U-algebras whose fields of quotients are galois extensions, it holds with respect to all finite U-algebras. We summarize the results of this discussion as follows:

PROPOSITION 1.1. *Let R be a local ring. If for each unramified R-algebra U with the same residue field as R, purity holds with respect to all finite U-algebras whose fields of quotients are galois extensions of the fields of quotients of U, then purity holds for R without any restrictions.*

Thus we lose no generality by only considering finite R-algebras S such that the field of quotients L of S is a galois extension of K with galois group G. We define the "twisted" group ring or trivial crossed product $S(G)$ as follows: its additive structure is that of a free left S-module with the elements of G as basis with multiplication defined by $(s_1g_1)(s_2g_2) = s_1(g_1(s_2))g_1g_2$. There is a natural ring homomorphism $\delta\colon S(G) \to \mathrm{Hom}_R(S,S)$ given by $\delta(sg)(x) = sg(x)$. It is clear that $S(G) \subset L(G)$ (where $L(G)$ is defined similarly) and that δ extends uniquely to a similar homomorphism $L(G) \to \mathrm{Hom}_K(L,L)$ which by classical galois theory is an isomorphism. Thus $\delta\colon S(G) \to \mathrm{Hom}_R(S,S)$ is a monomorphism, which becomes an isomorphism when tensored over R with K. Suppose now that every minimal prime ideal in S is unramified over R, or, what is the same thing, every minimal prime ideal in R is unramified over S. Now if $\mathfrak{p}$ is a minimal prime ideal in R, then $\delta_{\mathfrak{p}}\colon S_{\mathfrak{p}}(G) \to \mathrm{Hom}_{R_{\mathfrak{p}}}(S_{\mathfrak{p}}, S_{\mathfrak{p}})$ is an isomorphism since $S_{\mathfrak{p}}$ is an unramified $R_{\mathfrak{p}}$-algebra, which is a finitely generated, free $R_{\mathfrak{p}}$-module (see [5, prop. A.1]), Thus $R_{\mathfrak{p}} \otimes S(G) \to R_{\mathfrak{p}} \otimes \mathrm{Hom}_R(S,S)$ is an isomorphism for all minimal prime ideals in $\mathfrak{p}$. Also, since S is integrally closed, we know that S is a reflexive R-module, i.e., the natural map $h\colon S \to \mathrm{Hom}_R(\mathrm{Hom}_R(S,R),R)$ given by $h(s)(f) = f(s)$ is an isomorphism (see [4, Theorem 1.5]). Therefore $S(G)$ is R-reflexive since $S(G)$ is isomorphic to a direct sum of copies of S. Since $S(G)$ and $\mathrm{Hom}_R(S,S)$ span the same vector space and are isomorphic localized at the minimal primes, they are isomorphic (see [3, Proposition 3.4]). Thus we have that if the minimal primes of S are unramified, then $\mathrm{Hom}_R(S,S)$ is isomorphic to a direct sum of copies of S.

Suppose now R has the property that a reflexive module whose endomorphism ring is isomorphic (as an R-module) to a direct sum of copies of the module must be a free R-module. Then we could conclude that if all the minimal prime ideals of S are unramified, then S is R-free. But then S is unramified because a finite R-algebra which is a free R-algebra is unramified if its minimal primes are unramified (see [3]).

Thus we have established the following:

PROPOSITION 1.2. *Let R be a local ring such that every unramified R-algebra U with the same residue class field as R has the property that a reflexive U-module M is free if and only if* $\operatorname{Hom}_U(M,M)$ *is isomorphic to a direct sum of copies of M. Then purity holds in R.*

Suppose now that R is a regular local ring. Since every unramified R-algebra is also regular, to prove that purity holds for regular local rings it suffices, in view of Proposition 1.2. to prove

THEOREM 1.3. *If R is a regular local ring and M is a reflexive R-module such that* $\operatorname{Hom}_R(M,M)$ *is isomorphic to a direct sum of copies of M, then M is R-free.*

The rest of the paper is devoted to proving this theorem, as a result of which we obtain

THEOREM 1.4. *Purity holds for all regular local rings.*

2. Proof of theorem on modules. Before proceeding to the proof of Theorem 1.3, we give a summary of some of the homological facts about finitely generated modules over local rings used in the proof. (We shall always assume all modules are finitely generated).

If R is a local ring and M is a finitely generated R-module, then a sequence of non-units $x_1,\cdots,x_d$ in R is called an M-sequence if x_i is not a zero divisor in $M(x_1,\cdots,x_{i-1})M$ for $i=1,\cdots,d$. It can be shown that all M-sequences have length less than or equal to $\dim R$ and that all maximal M-sequences have the same length which is called the codimension of M (notation: $\operatorname{codim} M$). The primary decomposition theory of modules shows easily that an M-sequence $x_1,\cdots,x_d$ is maximal if and only if the maximal ideal $\mathfrak{m}$ in R belongs to zero in the module $M/(x_1,\cdots,x_d)M$. Again applying the primary decomposition theory it follows readily that an $\mathfrak{m}$-primary ideal $\mathfrak{a}$ always contains a maximal R-sequence. Now it is shown in [10] that the annihilator of a module M contains an R-sequence of length d if and only if $\operatorname{Ext}^i(M,R)=0$ for $i<d$. Thus it follows that if M is a module of finite length (i.e., the annihilator of M is $\mathfrak{m}$-primary) then $\operatorname{Ext}^i(M,R)=0$ for $i<\operatorname{codim} R$.

LEMMA 2.1. *Let $f\colon A\to B$ be a homomorphism of the R-modules A and B. If f induces an isomorphism $f_{\mathfrak{p}}\colon A_{\mathfrak{p}}\to B_{\mathfrak{p}}$ for all non-maximal prime ideals $\mathfrak{p}$ in R, then* $f^*\colon \operatorname{Ext}^i(B,R)\to\operatorname{Ext}^i(A,R)$ *is an isomorphism for* $i<(\operatorname{codim} R)-1$.

Proof. We have the exact sequences $0\to K\to A\to I\to 0$ and $0\to I\to B$

$\to C \to 0$ where $K = \operatorname{Ker} f$, $I = \operatorname{Im} f$ and $C = \operatorname{Coker} f$. Since $A_{\mathfrak{p}} \to B_{\mathfrak{p}}$ is an isomorphism for all non-maximal ideals $\mathfrak{p}$ in R, we know that K and C have finite length. From the usual exact sequences associated with the functor $\operatorname{Ext}(\ ,R)$ and the fact that $\operatorname{Ext}^i(K,R) = 0 = \operatorname{Ext}^i(C,R)$ for $i < \operatorname{codim} R$, we deduce that $\operatorname{Ext}^i(I,R) \approx \operatorname{Ext}^i(A,R)$ for $i < \operatorname{codim} R$, and $\operatorname{Ext}^i(B,R) \approx \operatorname{Ext}^i(I,C)$ for $i < (\operatorname{codim} R) - 1$, which gives the desored result.

The next proposition essentially outlines the method we shall follow in proving Theorem 1.3. Before stating the proposition we need a few definitions. We will call the R-module $\operatorname{Hom}(M,R)$ the dual of M and denote it by M^*. We always have a natural map $\phi\colon M \to M^{**}$ given by $\phi(m)(f) = f(m)$. We shall say that M is reflexive if $M \approx M^{**}$. It is well known that if M is R-projective, i.e., a free R-module, then M is reflexive.

PROPOSITION 2.2. *Let $\mathcal{S}$ be a set of pairs (R,M) where R is a local ring and M is a reflexive R-module, and suppose $\mathcal{S}$ satisfies the following conditions:*

a) *If (R,M) is in $\mathcal{S}$, then $(R_{\mathfrak{p}}, M_{\mathfrak{p}})$ is in $\mathcal{S}$ for each prime ideal $\mathfrak{p}$ in R.*

b) *If (R,M) is in $\mathcal{S}$ and $\operatorname{codim} R \leqq 3$, then M is R-free.*

c) *If (R,M) in $\mathcal{S}$ has the properties that $\operatorname{codim} R > 3$ and $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free for all non-maximal prime ideals p in R, then there is a non-unit x in R which is not a zero divisor for R and such that $(\bar{R}, \bar{M}^{**})$ is in $\mathcal{S}$ where $\bar{R} = R/(x)$ and $\bar{M} = M/xM$.*

Then for each pair (R,M) in $\mathcal{S}$, we have that M is R-free.

Proof. Suppose the proposition is false. Let (R,M) in $\mathcal{S}$ be a pair for which the proposition is false and such that R has the smallest dimension of all local rings occurring in pairs in $\mathcal{S}$ for which the proposition is false. Let x be a non-unit which is not a zero divisor for R and such that $(\bar{R}, \bar{M}^{**})$ is in $\mathcal{S}$. Since x is a non-unit which is not a zero divisor in R we know that $1 + \dim \bar{R} = \dim R$. Therefore $\bar{M}^{**}$ is $\bar{R}$-free. The fact that $\dim R_{\mathfrak{p}} < \dim R$ for each non-maximal prime ideal $\mathfrak{p}$ in R, gives us that $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free. Therefore $\bar{M}_{\bar{\mathfrak{p}}}$ is $\bar{R}_{\bar{\mathfrak{p}}}$-free for all non-maximal prime ideals $\bar{\mathfrak{p}}$ in $\bar{R}$. Thus it follows from Lemma 2.1, that $\operatorname{Ext}^i_{\bar{R}}(\bar{M},\bar{R}) \approx \operatorname{Ext}^i_{\bar{R}}(\bar{M}^{**},\bar{R})$ for $i < (\operatorname{codim} \bar{R}) - 1$. Now since M is by assumption not free, it follows that $\operatorname{codim} R \geqq 4$ and therefore $\operatorname{codim} \bar{R} \geqq 3$. Thus we have that $\operatorname{Ext}^i_{\bar{R}}(\bar{M},\bar{R}) \approx \operatorname{Ext}^i_{\bar{R}}(\bar{M}^{**},\bar{R})$ for $i \leqq 1$. Since $\bar{M}^{**}$ is $\bar{R}$-free we have that $0 = \operatorname{Ext}^i_{\bar{R}}(\bar{M}^{**},\bar{R}) = \operatorname{Ext}^i_{\bar{R}}(\bar{M},\bar{R})$.

Now the fact that x is not a zero divisor in R and M is reflexive, means that x is not a zero divisor in M. Therefore by a standard change of rings argument we have that $\operatorname{Ext}^i_R(M,\bar{R}) \approx \operatorname{Ext}^i_{\bar{R}}(\bar{M},\bar{R})$ for all i, which gives us that $\operatorname{Ext}^i_R(M,\bar{R})=0$. From the exact sequence $0\to R\xrightarrow{x}R\to\bar{R}\to 0$ we deduce the exact sequence

$$0\to M^*\xrightarrow{x}M^*\to\operatorname{Hom}_R(M,\bar{R})\to\operatorname{Ext}^1_R(M,R)$$
$$\xrightarrow{x}\operatorname{Ext}^1_R(M,R)\to\operatorname{Ext}^1_R(M,\bar{R}).$$

Since all the modules are finitely generated R-modules and $\operatorname{Ext}^1_R(M,\bar{R})=0$, we have by Nakayama's lemma that $\operatorname{Ext}^1_R(M,R)=0$. Also we have observed already that $\operatorname{Ext}^i_{\bar{R}}(\bar{M},\bar{R})\approx\operatorname{Ext}^i_{\bar{R}}(\bar{M}^{**},\bar{R})$ for $i\leqq 1$, and thus $\bar{M}^*=\bar{M}^{***}$ is a free $\bar{R}$-module. Since $\operatorname{Hom}_R(M,\bar{R})=\bar{M}^*$, we have that $M^*/xM^*=\bar{M}^*$ and therefore has homological dimension 1 over R. Thus M^* is R-free. But this contradicts the fact that (R,M) was chosen so that M was not free. This contradiction proves the proposition.

The rest of this note is devoted to showing that if $\mathcal{S}$ is the set of all pairs (R,M) where R is a regular local ring and M is a reflexive module such that $\operatorname{Hom}(M,M)$ is a direct sum of copies of M, then $\mathcal{S}$ satisfies the hypothesis of Proposition 2.2, which will give us the desired theorem.

The fact that $\mathcal{S}$ satisfies a) follows from the formal fact that $R_{\mathfrak{p}}\otimes_R\operatorname{Hom}(M,M)\approx\operatorname{Hom}_{R_{\mathfrak{p}}}(M_{\mathfrak{p}},M_{\mathfrak{p}})$ and the fact that if R is regular so is $R_{\mathfrak{p}}$ for all prime ideals $\mathfrak{p}$.

We next pass to showing that $\mathcal{S}$ satisfies condition b) which can be stated as follows: if R is a regular local ring whose dimension is less than or equal to 3 and M is a reflexive R-module such that $\operatorname{Hom}(M,M)$ is isomorphic to a direct sum of copies of M, then M is R-free. Since for a regular local ring of dimension less than or equal to 2, all reflexive modules are free (see [4] for instance), we can assume that $\dim R=3$ and $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free for all nonmaximal prime ideals p in R. Also, since M is reflexive, and $\operatorname{codim}R=3$, we have that $\operatorname{codim}M\geqq 2$. Since $\operatorname{codim}M+\operatorname{hd}M=\dim R$, we have that the $\operatorname{hd}M\leqq 1$.

Now if S is an arbitrary local ring and A is an S-module of finite homological dimension, we define the rank of A (notation: $\operatorname{rank}A$), as follows: let $0\to X_t\to\cdots\to X_1\to X_0\to A\to 0$ be exact with the X_i finitely generated free S-modules. Then if we denote by $r(X_i)$ the number of elements in a free basis of X_i, we define $\operatorname{rank}A=\sum(-1)^i r(X_i)$. Since $\sum(-1)^i r(X_i)$ $=\sum(-1)^i\dim_k(\operatorname{Tor}^S_i(A,k))$ where k is the residue class field of S, the number $\sum(-1)^i r(X_i)$ does not depend on the particular finite resolution chosen. Thus $\operatorname{rank}A_{\mathfrak{p}}=\operatorname{rank}A$ for any prime ideal $\mathfrak{p}$ in S. Now if $\mathfrak{p}$ is a minimal prime ideal in S, then $A_{\mathfrak{p}}$ is $S_{\mathfrak{p}}$-free since $\operatorname{codim}S_{\mathfrak{p}}=0$ and $\operatorname{hd}_{S_{\mathfrak{p}}}A_{\mathfrak{p}}$

$<\infty$. Thus $0\to (X_t)_{\mathfrak p}\to\cdot\cdot\cdot\to (X_0)_{\mathfrak p}\to A_{\mathfrak p}\to 0$ is an exact sequence of free free $S_{\mathfrak p}$-modules and therefore it follows that $r(A_{\mathfrak p}) = \sum(-1)^i r(X_i)_{\mathfrak p} = \operatorname{rank} A$. Thus the $\operatorname{rank} A \geqq 0$. It is now easy to see that if $\operatorname{Hom}(A,A)\approx\sum_{i=1}^{t} A_i$ where each $A_i\approx A$, then $t=\operatorname{rank} A$. Because this assumption is certainly true if A is free, and the general situation can be reduced to the case of A being free by tensoring with $S_{\mathfrak p}$ where $\mathfrak p$ is a minimal prime ideal in S.

Therefore returning to our original situation of R a regular local ring of dimension 3, M a reflexive R-module such that $\operatorname{Hom}(M,M)$ is isomorphic to a direct sum of copies of M, we have seen that $M_{\mathfrak p}$ is $R_{\mathfrak p}$-free for all non-maximal prime ideals $\mathfrak p$, that the $\operatorname{hd} M\leqq 1$ and that $\operatorname{Hom}(M,M)\approx\sum_{i=1}^{t} M_i$ where each $M_i\approx M$ and $t=\operatorname{rank} M$. The fact that M is R-free is now an immediate consequence of the following more general proposition.

PROPOSITION 2.3. *Let R be a local ring of codimension $\geqq 3$. Let M be a reflexive module such that*

a) *$M_{\mathfrak p}$ is $R_{\mathfrak p}$-free for all non-maximal prime ideals*

b) $\operatorname{hd} M\leqq 1$

c) *the modules* $\operatorname{Ext}^1(\operatorname{Hom}(M,M),R)$ *and* $\operatorname{Ext}^1(M,R)$ *which are of finite length have the property that*

$$L(\operatorname{Ext}^1(\operatorname{Hom}(M,M),R))\leqq(\operatorname{rank} M)L(\operatorname{Ext}^1(M,R)).$$

Then M is a free R-module.

Proof. We always have a natural homomorphism

$$\phi\colon M^*\otimes M\to\operatorname{Hom}(M,M)$$

given by $\phi(f\otimes m)(x)=f(x)m$ which has the property that it is an isomorphism if and only if M is R-free (see [4, Proposition A.1] for instance). Therefore since the $\operatorname{codim} R\geqq 3$, we have, applying Lemma 2.1, that

$$\operatorname{Ext}^1(M^*\otimes M,R)\approx\operatorname{Ext}^1(\operatorname{Hom}(M,M),R).$$

Also since $M_{\mathfrak p}$ is $R_{\mathfrak p}$-free for all non-maximal prime ideals $\mathfrak p$, we have that $R_{\mathfrak p}\otimes\operatorname{Ext}^1(M,R)=0=R_{\mathfrak p}\otimes\operatorname{Ext}^1(\operatorname{Hom}(M,M),R)$ for all non-maximal prime ideals $\mathfrak p$ in R and thus $\operatorname{Ext}^1(M,R)$ and $\operatorname{Ext}^1(\operatorname{Hom}(M,M),R)$ have finite length. Thus we have that $L(\operatorname{Ext}^1(M^*\otimes M,R))=L(\operatorname{Ext}^1(\operatorname{Hom}(M,M),R))$. We now concentrate our attention on $\operatorname{Ext}^1(M\otimes M^*,R)$.

Let $0\to X_1\to X_0\to M\to 0$ be exact with the X_i free. Then

$$(*)\qquad 0\to M^*\to X_0{}^*\to X_1{}^*\to\operatorname{Ext}^1(M,R)\to 0$$

is exact with the X^*_i free. Since the hd $M \leqq 1$, we know that

$$\operatorname{Tor}^R_i(M, \operatorname{Ext}^1(M, R)) = 0$$

for $i > 1$. Thus we know that $\operatorname{Tor}^R_i(M, M^*) = 0$ for $i > 0$ by the usual dimension shift type argument. Thus we have that

$$0 \to X_1 \otimes M^* \to X_0 \otimes M^* \to M \otimes M^* \to 0$$

is exact. Therefore $\operatorname{Hom}(X_1 \otimes M^*, R) \to \operatorname{Hom}(X_0 \otimes M^*) \to \operatorname{Ext}^1(M \otimes M^*, R)$ is exact. But by the usual associativity formulas we have that

$$\operatorname{Hom}(X_i \otimes M^*, R) = \operatorname{Hom}(X_i, M^{**}).$$

Observing that $M = M^{**}$ and making the necessary substitutions, we have the exact sequence $\operatorname{Hom}(X_1, M) \to \operatorname{Hom}(X_0, M) \to \operatorname{Ext}^1(M \otimes M^*, R)$. But the $\operatorname{Coker}(\operatorname{Hom}(X_1, M) \to \operatorname{Hom}(X_0, M)) = \operatorname{Ext}^1(M, M)$. Thus we have an exact sequence $0 \to \operatorname{Ext}^1(M, M) \to \operatorname{Ext}^1(M \otimes M^*, R)$.[2] Therefore if we show that assuming that the hd $M = 1$ implies that

$$L(\operatorname{Ext}^1(M, M)) > (\operatorname{rank} M) L(\operatorname{Ext}^1(M, R)),$$

we will be done, since

$$L(\operatorname{Ext}^1(\operatorname{Hom}(M, M), R)) = L(\operatorname{Ext}^1(M \otimes M^*, R)) \geqq L(\operatorname{Ext}^1(M, M)).$$

Tensoring the exact sequence (*) with M we have the exact sequence $X_0^* \otimes M \to X_1^* \otimes M \to \operatorname{Ext}^1(M, R) \otimes M \to 0$. Now we have natural maps $\beta_i \colon X_i^* \otimes M \to \operatorname{Hom}(X, M)$ defined by $\beta_i(f \otimes m)(x) = f(x)m$ which are easily seen to be isomorphisms since the X_i are finitely generated projective R-modules. Therefore we have the exact sequence $\operatorname{Hom}(X_0, M) \to \operatorname{Hom}(X_1, M) \to \operatorname{Ext}^1(M, R) \otimes M \to 0$. But the $\operatorname{Coker}(\operatorname{Hom}(X_0, M) \to \operatorname{Hom}(X_1, M)) = \operatorname{Ext}^1(M, M)$. Therefore we have that $\operatorname{Ext}^1(M, R) \otimes M \approx \operatorname{Ext}^1(M, M)$.[3]

Assume hd $M = 1$. Then $\operatorname{Ext}^1(M, R) \neq 0$ (see [4]), and is of finite length. Let $C = \operatorname{Ext}^1(M, R)$. Then from the exact sequence $0 \to \operatorname{Tor}^R_1(C, M) \to C \otimes X_1 \to C \otimes X_0 \to C \otimes M \to 0$ we conclude that

$$\begin{aligned} L(C \otimes M) &= L(X_0 \otimes C) - L(X_1 \otimes C) + L(\operatorname{Tor}^R_1(C, M)) \\ &= (\operatorname{rank} M) L(C) + L(\operatorname{Tor}^R_1(C, M)). \end{aligned}$$

---

[2] This is a special case of a longer exact sequence associated with a standard spectral sequence.

[3] This is a special case of the following more general theorem: let Λ be a left noetherian ring (not necessarily commutative) and A a left (finitely generated) Λ-module such that the $\operatorname{hd}_\Lambda A = j < \infty$. Then for each left Λ-module C we have that $\operatorname{Ext}^j(A, C) \approx \operatorname{Ext}^j(A, \Lambda) \otimes_\Lambda C$ where $\operatorname{Ext}^j(A, \Lambda)$ is considered a right Λ-module by the right operations of Λ on itself. The proof is essentially the same as that given above. Similarly one can also show that $\operatorname{Tor}^j(B, A) = \operatorname{Hom}(\operatorname{Ext}^j(A, \Lambda), B)$ for all right Λ-modules B.

But since $C \neq 0$ and has finite length and $\operatorname{hd} M = 1$, we know by [2, Proposition 1.1] that $\operatorname{Tor}^R{}_1(C, M) \neq 0$. Thus

$$L(\operatorname{Ext}^1(M,M)) = L(\operatorname{Ext}^1(M,R) \otimes M) > (\operatorname{rank} M) L(\operatorname{Ext}^1(M,R)),$$

which completes the proof.

We now show that the set $\mathcal{S}$ consisting of all pairs (R, M) with R regular and M a reflexive R-module such that $\operatorname{Hom}(M,M)$ is isomorphic to a direct sum of copies of M satisfies condition c) of Proposition 2.2.

Let (R, M) be in $\mathcal{S}$ such that $\dim R > 3$ and $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free for all non-maximal prime ideals $\mathfrak{p}$ in R. Let x be in $\mathfrak{m} - \mathfrak{m}^2$ where $\mathfrak{m}$ is the maximal ideal of R. Since M is reflexive, the element x is not a zero divisor in M. From the exact sequence $0 \to M \xrightarrow{x} M \to \bar{M} \to 0$ (where $\bar{M} = M/xM$) we deduce the exact sequence

$$0 \to \operatorname{Hom}(M,M) \xrightarrow{x} \operatorname{Hom}(M,M) \to \operatorname{Hom}_R(M, \bar{M}) \to \operatorname{Ext}^1(M,M).$$

Thus we get an exact sequence of $\bar{R} = R/(x)$-modules

$$0 \to \sum \bar{M}_i \to \operatorname{Hom}_{\bar{R}}(\bar{M}, \bar{M}) \to T \to 0$$

where $\sum \bar{M}_i$ is a direct sum of modules isomorphic to $\bar{M}$ and T has finite length (since $M_{\mathfrak{p}}$ is $R_{\mathfrak{p}}$-free for all $\mathfrak{p} \neq \mathfrak{m}$, we know that $\operatorname{Ext}^1(M,M)$ has finite length). Also we know that $\bar{R}$ is a regular local ring of dimension greater than or equal to 3. Therefore by Lemma 2.1 we have that

$$\operatorname{Hom}_{\bar{R}}(\bar{M}, \bar{M})^* \approx \sum \bar{M}^*{}_i \text{ (direct sum)}$$

since $T_{\mathfrak{p}} = 0$ for all non-maximal prime ideals $\mathfrak{p}$ in $\bar{R}$. Thus $\operatorname{Hom}_{\bar{R}}(\bar{M}, \bar{M})^{**} = \sum M^{**}{}_i$. But since $\bar{R}$ is regular and thus integrally closed, we have by [4, Proposition 4.6], that $\operatorname{Hom}_{\bar{R}}(\bar{M}, \bar{M})^{**} = \operatorname{Hom}_{\bar{R}}(\bar{M}^{**}, \bar{M}^{**})$. Therefore $(\bar{R}, \bar{M}^{**})$ is in $\mathcal{S}$ which shows that $\mathcal{S}$ satisfies all the conditions of Proposition 2.2. Therefore if R is a regular local ring and M is a reflexive R-module such that $\operatorname{Hom}(M,M)$ is isomorphic with a direct sum of copies of M, then M is free, which establishes Theorem 1.3.

Remark. The procedure outlined in Proposition 2.2 for proving that a module over a regular local ring is free was employed [4] to show that if a reflexive module has a free module as endomorphism ring, then the module is free. In that case the set $\mathcal{S}$ was taken to be all pairs (R, M) with R a regular local ring and M a reflexive module such that $\operatorname{Hom}(M,M)$ is a free R-module. It should be observed that Proposition 2.3 of this paper shows that this $\mathcal{S}$

satisfies condition b) of Proposition 2.2. Since condition a) is clearly satisfied, it remains only to show that c) is satisfied, for which we refer the reader to [4].

It would be interesting to know if Theorem 1.3 of this paper and the above-stated theorem in [4] are valid if instead of assuming that R is regular, we assume instead that the homological dimension of M is finite. Again condition a) of Proposition 2.2 is trivial and condition b) is shown to hold by Proposition 2.3. The only difficulty in showing that condition c) holds is that it is not clear that $\bar{M}^{**}$ (in the notation of Proposition 2.2) has finite homological dimension.

Collège de France,
Brandeis University.

REFERENCES.

[1] S. Abhyankar, *Ramification Theoretic Methods in Algebraic Geometry*, Annals of Mathematics Studies No. 43, Princeton University.

[2] M. Auslander, "Modules over unramified regular local rings," *Illinois Journal of Mathematics* (in press).

[3] M. Auslander and D. A. Buchsbaum, " On ramification theory in noetherian rings," *American Journal of Mathematics*, vol. 81 (1959), pp. 749-765.

[4] M. Auslander and O. Goldman, "Maximal orders," *Transactions of the American Mathematical Society*, vol. 97 (1960), pp. 1-24.

[5] ———, "The Brauer group of a commutative ring," *Transactions of the American Mathematical Society*, vol. 97 (1960), pp. 367-409.

[6] M. Auslander and D. S. Rim, "Ramification index and multiplicity," (to appear).

[7] W. Chow, "On the theorem of Bertini for local domains," *Proceedings of the National Academy of Science, U.S.A.*, vol. 44 (1958), pp. 580-584.

[8] A. Grothendieck, "Morphisms étales-*I*," *Seminare de Géometrie Algébrique de l'Institut des Hautes Études Scientifique* (1960) (mimeographed lecture notes).

[9] M. Nagata, "Purity of branch loci in regular local rings," *Illinois Journal of Mathematics*, vol. 3 (1959), pp. 328-333.

[10] D. Rees, "The grade of an ideal or module," *Proceedings of the Cambridge Philosophical Society*, vol. 53 (1957), pp. 28-42.

[11] O. Zariski, " On the purity of the branch locus of algebraic functions," *Proceedings of the National Academy of Science, U.S.A.*, vol. 44 (1958), pp. 791-796.

ON THE NUMBER OF LATTICE POINTS IN A CONVEX SET.*[1]

By C. S. Herz.

---

0. Statement of results. Let C be a compact subset of euclidean space E^{n+1}. We denote by $\Lambda(C)$ the number of lattice points in C and by $V(C)$ the volume of C. Our aim is to estimate $\Delta(C) = \Lambda(C) - V(C)$. In this paper we obtain crude bounds for Δ which depend only on euclidean invariants by means of a short and elementary deduction from some facts of Fourier analysis. The results are stated in a form suitable for application when C is a suitably well-behaved convex body.

The main idea involves consideration of the Fourier transform of the characteristic function of C, i. e. the function ϕ defined by

$$\phi(C,t) = \int_C \exp(2\pi i t \cdot x)\,dx \tag{1}$$

Under appropriate hypotheses on the set C we shall have $\phi(C,t) = O(|t|^{-\nu-1})$ as $|t| \to \infty$ where $\nu = n/2$. This motivates the introduction of a quantity ρ defined by

$$\rho^{\nu}(C) = c_n^{-1} \sup_t |t|^{\nu+1} |\phi(C,t)|, \tag{2}$$

c_n being some convenient normalizing constant. It is evident from the definition that ρ is a euclidean invariant. Moreover ρ is homogeneous of degree 1, that is to say, if $h > 0$ and hC is the homothetic dilation of C we have $\rho(hC) = h\rho(C)$. This last fact follows directly from the obvious identity: $\phi(hC,t) = h^{n+1}\phi(C,ht)$.

The point of the preceding paragraph is that an inequality of the form

$$|\phi(C,t)| \leqq c_n \rho^{\nu}(C)\,|t|^{-\nu-1} \tag{2'}$$

is available with ρ a euclidean invariant homogeneous of degree 1; to be sure, we may have $\rho(C) = +\infty$ for some C, but that is not relevant to the immediate conclusion which is

---

* Received October 20, 1961.

[1] The research for this paper was sponsored by the National Science Foundation, Contract No. G 18837.

THEOREM 1. *Suppose C is a convex compact subset of E^{n+1} If $V(C) \geqq 1$ then*

$$|\Delta(C)| \leqq k_n \rho^{n\alpha_n}(C) \tag{3}$$

where k_n depends only on n and $\alpha_n = (n+1)(n+2)^{-1}$.

What we want to emphasize is that (3) is a consequence of (2′). In order that the theorem be useful we have either to provide a more explicit description of $\rho(C)$ or list conditions under which $\rho(C)$ is finite. The decisive role is played by the boundary, $\dot{C}$, of C, and Theorem 1 becomes more enlightening after two remarks.

Remark A (Herz, [3]). *If $\dot{C}$ is of differentiability class $C^{(m)}$, $m \geqq [\nu + 2]$,[2] and has strictly positive gaussian curvature, then $\rho(C)$ is finite.*

Remark B (van der Corput, [1]). *In case $n = 1$ we may take $\rho(C)$ to be the maximum radius of curvature of $\dot{C}$.*

The statements embodied in the two remarks are both easily proved by straightforward integration by parts in well-chosen coordinate systems. Nevertheless, even Remark A has rather striking consequences. For dilations of a fixed convex body we have the trivial estimate: $\Delta(hC) = O(h^n)$ *as* $h \to \infty$. We may conclude from Theorem 1 that if the hypotheses of Remark A hold then $\Delta(hC) = O(h^{n\alpha_n})$. Now, while it seems highly likely that with further differentiability assumptions on C one can obtain $O(h^{n\alpha_n - \epsilon})$ where $\epsilon > 0$, the estimates have been improved only slightly[3] even for balls centered at the origin when $n = 1$ or 2; and the improvements have come at the expense of considerable refinement of technique involving careful estimation of trigonometric sums. (When $n = 1$, van der Corput, [2], has shown that if $\dot{C}$ is of class $C^{(5)}$ then $\Delta(hC) = O(h^{2/3-\epsilon})$ with $\epsilon > 0$.) Thus Theorem 1 in conjunction with Remark A can be regarded as a reasonably satisfactory result as far as dilations of a fixed set are concerned. The general statement, however, suffers from the fact that ρ is not specified in geometric terms when $n > 1$. In case $n = 1$, the coupling of Remark B with Theorem 1 leads to a best possible result, cf. [4], which we should like to imitate in higher dimensions. Theorem 1 itself already suggests the form we should seek to obtain.

---

[2] The square brackets denote "greatest integer in."

[3] The best results to date are due to Hua, *Quarterly Journ.* 13 (1942), pp. 18-29, and to Vinogradov, *Izvestia Academia Nauk* 19 (1955), pp. 3-10, who have extended the permissible upper bounds on ϵ in the cases $n = 1$ and $n = 2$ respectively to $\epsilon < 1/60$ and $\epsilon < 1/8$.

If the body C is obtained from a body B as a result of an affine transformation A then we easily see from (2) that

$$\rho^{\nu}(C) \leqq a^{-\nu-1} \,|\det A\,|\, \rho^{\nu}(B)$$

where a is the smallest eigenvalue of $|A|$. From this, we can derive the result that when $\dot{C}$ is an ellipsoid, $\rho^{n}(C) \leqq b_n K^{-1}(C)$, where $K(C)$ is the minimum of the gaussian curvature of $\dot{C}$. Theorem 1 now gives rise to

COROLLARY. *Let* $\dot{C}$ *be an ellipsoid in* E^{n+1}. *If* $V(C) \geqq 1$ *then*

$$|\Delta(C)| \leqq k_n K^{-\alpha_n}(C)$$

where $K(C)$ *is the minimum of the gaussian curvature of* $\dot{C}$.

The corollary would appear to be a perfect generalization of van der Corput's Theorem, i.e. Theorem 1 in case $n=1$ and Remark B. Unfortunately, it seems doubtful that this generalization is valid for general convex bodies. The higher derivatives of surface quantities do seem to play a role when $n>1$. Something can be salvaged, however. The asymptotic formula of [3, Theorem 3] shows that the inequality (2′) can be replaced by

$$(4) \qquad |\phi(C,t)| \leqq \pi^{-1} K^{-\frac{1}{2}}(C)\,|t|^{-\nu-1} + L(C)\,|t|^{-\nu-2}$$

where L is a euclidean invariant homogeneous of degree $\nu-1$ which is finite under the hypotheses of Remark A with $\nu+2$ replaced by $\nu+4$. A slightly more sophisticated version of the reasoning leading from (2′) to Theorem 1 leads from (4) to Theorem 2 below. Our methods of proof, however, seem to require a hypothesis that C is sufficiently thick which we state in terms of a condition involving the diameter $\delta(C)$.

THEOREM 2. *Let* C *be a convex subset of* E^{n+1}. *If* $V^2(C) \geqq \delta^{n+2}(C) \geqq 1$ *then*

$$(5) \qquad |\Delta(C)| \leqq k'_n K^{-\alpha_n}(C) + E(C)$$

where k'_n *depends only on* n *and* E *is a euclidean invariant homogeneous of degree* $(n-1)\alpha_n$ *which is finite whenever* C *is of differentiability class* $C^{(m)}$, $m \geqq [\nu+4]$, *and has positive minimum gaussian curvature,* $K(C)$. The assertion that E is homogeneous of degree $(n-1)\alpha_n$ is valid only for $n>2$; when $n=2$ it should be replaced by $E(hC)=O(\log h)$ as $h\to\infty$; only the cases $n>1$ are of interest here.

In estimate (5) the term containing the gaussian curvature is homogeneous of degree $n\alpha_n$. The theorem is supposed to suggest that if C is "big" then the error term is negligible in comparison with the principal term.

1. The first method of estimation. Let C be an arbitrary compact set in E^{n+1}. If D is another compact set, the convolution of the characteristic function of $C+D$ (vector sum of sets) and the characteristic function of $-D$ is non-negative and has the value $V(D)$ at each point of C. The Fourier transform of this convolution is the product $\phi(C+D,t)\bar{\phi}(D,t)$ where the bar denotes complex conjugate. Utilizing the Poisson summation formula, we conclude that

$$V(D)\Delta(C) \leqq \sum \phi(C+D,t)\bar{\phi}(D,t) \tag{6}$$

where the sum is extended over all lattice points t. Since $\phi(D,0)=V(D)$ and $\phi(C+D,0)=V(C+D)$, formula (6) can be rewritten as

$$\Delta(C) \leqq V(C+D) - V(C) + \sum{}' \phi(C+D,t)V^{-1}(D)\bar{\phi}(D,t) \tag{6'}$$

where $\sum'$ denotes a sum over all lattice points except the origin.

Now suppose that C is a convex body and that $D=hC$, the homothetic image of C by a factor $h>0$. On account of the convexity, $C+D=gC$ where $g=1+h$. Direct substitution in (6′) yields

$$\Delta(C) \leqq (g^{n+1}-1)V(C) + \sum{}' \phi(gC,t)V^{-1}(hC)\bar{\phi}(hC,t), \tag{7}$$

where $h>0$ and $g=1+h$.

The point of bothering to write down (7) is that we also have

$$\Delta(C) \geqq -(1-g^{n+1})V(C) + \sum{}' \phi(gC,t)V^{-1}(hC)\phi(hC,t), \tag{7'}$$

where $0<h<1$ and $g=1-h$. This formula is obtained by reasoning that, in the case of a compact convex body, the convolution of the characteristic functions of $(1-h)C$ and hC vanishes outside of C and is bounded above by $V(hC)$. Thenceforth the deduction is strictly analogous to that leading to (7).

Here we are only seeking a crude estimate,—one in which the oscillatory nature of $\phi(C,t)$ is not utilized. We are thus content to replace an estimate like (6) with

$$\Delta(C) \leqq V(C+D) - V(C) + \sum{}' |\phi(C+D,t)|\,|V^{-1}(D)\phi(D,t)|.$$

In doing this we make the bound independent of translations of C, i.e. we have a simultaneous upper bound on $\Delta(C+x)$ for all x. For all t, $|V^{-1}(D)\phi(D,t)| \leqq 1$; this will be a reasonably good estimate for small t, say $|t| \leqq T$. In the complementary range, $|t|>T$, we use the Schwarz inequality on the sum. The net result is to replace (6′) by

$$\begin{aligned}\Delta(C) \leqq V(C+D) - V(C) &+ \Phi(C+D,T) \\ &+ \Psi(C+D,T)V^{-1}(D)\Psi(D,T)\end{aligned} \tag{8}$$

where Φ and Ψ are defined by

$$\Phi(C,T) = \sum_{1 \leqq |t| \leqq T} |\phi(C,t)| \tag{9}$$

and

$$\Psi^2(C,T) = \sum_{|t|>T} |\phi(C,t)|^2. \tag{10}$$

2. Proof of Theorem 1. Estimate (2) implies the inequalities and

$$\Phi(C,T) \leqq \text{const.}\, \rho^{\nu}(C) T^{\nu} \tag{11}$$

and

$$\Psi(C,T) \leqq \text{const.}\, \rho^{\nu}(C) T^{-\frac{1}{2}}. \tag{12}$$

The reasoning which led to (8) may be applied to (7) and (7′) to yield, for $0 < h < 1$, $T > 1$

$$\begin{aligned} |\Delta(C)| \leqq \text{const.}\, hV(C) &+ \text{const.}\, \rho^{\nu}(C) T^{\nu} \\ &+ \text{const.}\, h^{-\nu-1} \rho^{2\nu}(C) V^{-1}(C) T^{-1}. \end{aligned} \tag{13}$$

The optimal choices are $h = V^{-2/(n+2)}$ and $T = \rho^{n/(n+2)}$. With these we have $|\Delta(C)| \leqq \text{const.}\, V^{n(n+2)}(C) + \text{const.}\, \rho^{n\alpha}(C)$ and thence (3) via an isoperimetric inequality: $V \leqq \text{const.}\, \rho^{n+1}$. This last inequality is an elementary deduction from three facts: inequality (2), $|\phi(C,t)| \leqq V(C)$, and $V(C) = \int |\phi(C,t)|^2\, dt$. We have to have $V(C) \geqq 1$ in order to ensure that $h \leqq 1$.

3. Proof of Theorem 2. If D is any convex compact set such that 0 is the only lattice point interior to $D - D$, in particular if $\delta(D) \leqq 1$, then according to [3, Theorem 2] we have

$$\sum_{|t| \leqq T} |t|^2 |\phi(D,t)|^2 \leqq A_n S(D) T$$

where A_n is a dimensional constant and $S(D)$ is the surface area of D. A direct consequence of this is

$$\Psi^2(D,T) < A_n S(D) T^{-1}. \tag{14}$$

If we make use of (4) the resulting estimates are

$$\Phi(C,T) \leqq \text{const.}\, K^{-\frac{1}{2}}(C) T^{\nu} + \text{const.}\, L(C) T^{\nu-1} \tag{15}$$

($T^{\nu-1}$ should be replaced by $\log T$ when $n=2$) and

$$\Psi(C,T) \leqq \text{const.}\, K^{-\frac{1}{2}}(C)T^{-\frac{1}{2}} + \text{const.}\, L(C)T^{-\frac{3}{2}}. \tag{16}$$

Now, assuming that $h < \delta^{-1}(C)$, we can use (14), (15), and (16) to estimate (7) and (7′). The end result is

$$\begin{aligned} |\Delta(C)| \leqq \text{const.}\, hV(C) & \\ + K^{-\frac{1}{2}}(C) & \leqq \{\text{const.}\, T^{\nu} + \text{const.}\, h^{-\nu-1}S^{\frac{1}{2}}V^{-1}T^{-1}\} \\ & + L(C)\{\text{const.}\, T^{\nu-1} + \text{const.}\, h^{-\nu-1}S^{\frac{1}{2}}V^{-1}T^{-2}\}. \end{aligned} \tag{17}$$

It is appropriate to choose $T = S^{1/(n+2)}$ and $h = V^{-2/(n+2)}$ whence (5) follows in short order.

4. An alternative method of estimation. The idea used in Section 1 to establish (6) can be replaced by a slightly different approach which we shall sketch here. It is obvious that $C \subset C + D + x$ whenever $x \epsilon -D$; so we have

$$\Delta(C) \leqq \inf_{x \epsilon -D} \Delta(C + D + x). \tag{18}$$

A trivial rewriting furnishes the inequality

$$\Delta(C) \leqq V(C+D) - V(C) + \inf_{x\epsilon -D} \Delta(C+D+x). \tag{18′}$$

The idea now is to observe that $\Delta(C+x)$ cannot be too large at each point of a sufficiently large set. To see this write $\Delta = \Delta_1 + \Delta_2$ where

$$\Delta_1(C+x) \sim \sum_{|t| \leqq T} \phi(C,t) \exp(2\pi i t \cdot x)$$

and

$$\Delta_2(C+x) \sim \sum_{|t| > T} \phi(C,t) \exp(2\pi i t \cdot x).$$

We obviously have that

$$|\Delta_1(C+x)| \leqq \Phi(C,T) \tag{19}$$

and

$$\int |\Delta_2(C+x)|^2 \, dx = \Psi^2(C,T). \tag{20}$$

The integration in x is over a fundamental period. At this place we use the reasoning of the Tchebychev inequality: Given $\epsilon > 0$, in each subset of the fundamental period which is of measure at least ϵ there must be a point x

where $|\Delta_2(C+x)| \leqq \epsilon^{-\frac{1}{2}}\Psi(C,T)$. Suppose now that $-D$ is contained in a fundamental period which is certainly true if $\delta(D) \leqq 1$. We shall then have for some $x \epsilon -D$

$$\Delta_2(C+D+x) \leqq V^{-\frac{1}{2}}(D)\Psi(C+D,T). \tag{21}$$

The estimate (19), for $C+D$ instead of C, and (21) combine to furnish

$$\Delta(C) \leqq V(C+D) - V(C) + \Phi(C+D,T) + V^{-\frac{1}{2}}(D)\Psi(C+D,T). \tag{22}$$

This inequality should be compared with (8).

The reasoning can now be continued as before using (11) and (12) or (15) and (16). We take $D = hC$ and $h = V^{-2/(n+2)}$ as before, but now choose $T = V^{n/(n+1)(n+2)}$. The estimates obtained this way are slightly better than those of the previous method, but here we require $\delta(D) \leqq 1$ as in the proof of Theorem 2. The method of this section leads to a proof of Theorem 2 with (5) replaced by the stronger inequality

$$|\Delta(C)| \leqq k'_n K^{-\frac{1}{2}}(C)\, V^{\beta_n}(C) + k''_n L(C)\, V^{\gamma_n}(C) \tag{23}$$

where $2\beta_n = n^2(n+1)^{-1}(n+2)^{-1}$ and $2\gamma_n = n(n-2)(n+1)^{-1}(n+2)^{-1}$.

CORNELL UNIVERSITY.

REFERENCES.

[1] J. G. van der Corput, "Uber Gitterpunkte in der Ebene," ***Mathematische Annalen***, vol. 81 (1920), pp. 1-20.

[2] ———, "Neue Zahlentheoret ische Abschatzungen," ***Mathematische Annalen***, vol. 89 (1923), pp. 215-254.

[3] C. S. Herz, "Fourier transforms related to convex sets," ***Annals of Mathematics***, vol. 75 (1962), pp. 81-92.

[4] V. Jarnik, "Uber die Gitterpunkte auf konvexen Kurven," ***Mathematische Zeitschrift***, vol. 24 (1926), pp. 500-518.

Addendum (*added in proof*).

The author has just learned, in conversation with Dr. H. J. Reiter, of the interesting work of Edmund Hlawka, "Integrale auf konvexen Körpern," which appeared in *Monatshefte für Mathematik*, vol. 54 (1950), pp. 1-36, 81-99. He wishes to acknowledge Hlawka's priority for some of the results stated here.

Theorem 1, coupled with Remark A, is substantially contained in Satz 9 on page 25 of the first part of Hlawka's paper. The methods of proof differ.

Remark A and formula (4) above are stated to be consequences of [3, Theorem 3]. That theorem, insofar as it deals with the Fourier transform of the volume element of a convex set occurs, with stronger smoothness hypotheses, as Satz 1 on page 82 of the second part of Hlawka's paper. The weaker hypotheses in [3] are an improvement of little interest, it would seem; and, in addition, the part of [3, Theorem 3] concerning the Fourier transform of the surface element is implicit in Hlawka's work. Therefore, the main theorem of [3] should be viewed as a minor emendation of Hlawka's theorem.

APPROXIMATION BY SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS.*

By Felix E. Browder.

Introduction. Let G be an open subset of E^n, G_1 an open subset of G with compact closure in G. Suppose that A is a partial operator on G. We consider families of solutions S of $Au = 0$ in G, S_1 of $Au_1 = 0$ in G_1. If we give a norm on S_1, we may ask the question: When can every u_1 in S_1 be approximated with arbitrary precision in the given norm by functions u in S? As a special case of this question, let A be the Cauchy-Riemann operator in the plane E^2, S the family of holomorphic functions in the plane, S_1 the family of holomorphic functions in the open subset G_1 continuous in $\bar{G}_1$, and the norm the uniform norm on G_1. We obtain then the classical setting for the uniform approximation theorems associated with the names of Walsh, Lavrentiev, and Mergelyan ([12]).

In the present paper, we construct a general framework for dealing with this question for various classes of differential operators and normed function spaces. In particular, in Section 3 we obtain a generalization of the Walsh theorem for general linear elliptic operators for which uniqueness in the Cauchy problem holds. A generalized Runge's theorem for such equations has been established by Lax [9] and Malgrange [11]. Some generalizations of Mergelyan's Theorem were given previously by the writer in [3] as a part of a general body of related approximation theorems, but not the proof of the Walsh Theorem in the general case. The analytical apparatus of the present paper is stronger than that of [3] and rests essentially upon Lemma 8 below. In Section 6, we give an extension of our approximation theory to hyperbolic operators, using the results of Leray [11], and to operators of principal normal type having strongly pseudo-convex level surfaces, using the definitions and results of Hörmander [8].

Sections 1 and 2 present a general formulation and solution of the approximation problem under abstract hypotheses. Section 3 applies these results to the proof of the generalized Walsh theorem. Section 4 considers a

---

* Received October 25, 1961.

related result concerning approximation by linear combinations of fundamental solutions with poles in a prescribed exterior domain. Section 5 considers approximations by solutions of $Au = 0$ with prescribed boundary values on a portion of the boundary of the larger domain, generalizing some weaker results of H. Beckert [1] for second-order equations. Section 6 gives the extension to the treatment of non-elliptic operators.

The writer should like to thank S. Agmon for suggesting the question treated in Section 4, and A. P. Calderon for describing to the writer his proof of Lemma 9.

Section 1. Let G be an open subset of Euclidean n-space E^n. The points of E^n will be denoted by $x = (x_1, \cdots, x_n)$, D_j will denote the elementary differential operator $i^{-1}\dfrac{\partial}{\partial x_j}$ for $1 \leqq j \leqq n$, and $|x| = \sqrt{\sum x^2{}_j}$ the Euclidean length in E^n. For each n-vector $\alpha = (\alpha_1, \cdots, \alpha_n)$ of non-negative integers, we set $|\alpha| = \sum_{j=1}^{n} \alpha_j$ and

$$D^\alpha = \prod_{j=1}^{n} D_j^{\alpha_j}.$$

Let A be a linear differential operator of order r with coefficients defined on G. Then A may be written in the form

$$A = \sum_{|\alpha| \leqq r} a_\alpha(x) D^\alpha,$$

where the coefficients $a_\alpha(x)$ are assumed to be continuous complex-valued functions defined on G. We define the formal adjoint differential operator A' for A by the prescription

$$A'u = \sum_{|\alpha| \leqq r} D^\alpha(\bar{a}_\alpha u) = \sum_{|\alpha| \leqq r} a'_\alpha D^\alpha u,$$

where the coefficients a'_α in the second expression on the right are distributions obtained by expanding the derivatives in the middle expression by the Leibniz formula.

Let G_1 be an open subset of G having compact closure in G. Suppose that we are given a locally convex linear topological vector space F, over the complex field C, of distributions (or possibly functions) on G such that F contains $C^\infty{}_c(G)$, the space of infinitely differentiable functions with compact support in G, while the imbedding maps of $C^\infty{}_c(G)$ into F and of F into $\mathfrak{D}'(G)$, the space of distributions on G, are both continuous. Let F^* be the dual space of F, (f, f^*) the pairing between an element f of F and f^* of the dual space F^*. (We remark that F^* is defined to be the space of continuous conjugate-linear functionals on F, rather than the linear functionals on F, so that $(f, f^*) = \bar{f}^*(f)$ is a sesqui-linear form, i.e. linear in f and conjugate linear in f^*.)

We shall assume that $C^\infty{}_0(G)$ is contained and weak* dense in F^*, that F^* is contained in $\mathcal{D}'(G)$, and that the pairing (f, f^*) between F and F^* is an extension by continuity of the pairing between f as a distribution on G and f^* restricted to lie in $C^\infty{}_0(G)$. Similarly we may consider a space F_1 of distributions on G_1 and its dual space $F_1{}^*$ and again denote the pairing between these two spaces by $(f_1, f_1{}^*)$. Between these pairs of spaces, we require the following relation:

$(\mathrm{a})_0$ *The restriction of F to the subset G_1 of G (i.e. the set of $f|_{G_1}$ for f in F) is contained in F and dense in F_1. The restriction map of F into F_1 is continuous.*

In terms of the spaces F and F_1, we may formulate our approximation problem in the following terms:

Let

$$S \equiv \{f : f \in F, Af = 0 \text{ on } G\},$$
$$S_1 \equiv \{f_1 : f_1 \in F_1, Af_1 = 0 \text{ on } G_1\}.$$

Consider the closure of the set $S|_{G_1}$ in F_1, i.e. the closure of the set of $f|_{G_1}$ for f in S.

When is this closure (which is always contained in S_1) equal to S_1?

To make this formulation more precise, we must make more definite the meaning of the equation $Af = 0$ in G for an element f of F. To do this, we shall make the following regularity assumption upon the coefficients of the differential operator A':

For each ψ in $C^\infty{}_0(G)$, $A'\psi$ must lie in F^*.

(It suffices for example to assume that all the coefficients a'_α of A' lie in F^* themselves.) Under this assumption, which we shall make without explicit mention throughout our discussion, we may define the maximal operator T of A in F by

$$D(T) = \{f : f \in F, \text{ There exists } g \text{ in } F \text{ such that } (f, A'\psi) = (g, \psi) \text{ for all } \psi \text{ in } C^\infty{}_0(G)\},$$
$$Tf = g.$$

With the definition of the maximal operator T of A in F, and the parallel definition of the maximal operator of A in F_1 on the open subset G_1, we set $S = N(T)$, the nullspace of T, and $S_1 = N(T_1)$, the nullspace of T_1.

Let us denote by J the operator which maps F into F_1 by mapping each f in F into $f|_{G_1}$. Its adjoint map $J_1 = J^*$ maps $F_1{}^*$ injectively in F^* by setting

$$(f, J_1(f_1{}^*)) = (f|_{G_1}, f_1{}^*).$$

J_1, considered as acting on the elements of $F_1{}^*$ as distributions, sets them equal to zero outside of G_1. Hence $J_1(f_1{}^*)$ for each $f_1{}^*$ has its support in the closure of G_1.

We can characterize the closure of $S|_{G_1}$ in F_1, i. e. of $J(S)$, by characterizing its annihilator $U = (J(S))^{\perp}$ in $F_1{}^*$. An element $f_1{}^*$ of $F_1{}^*$ will belong to U iff $(J(f), f_1{}^*) = 0$ for all f in S, or equivalently iff $(f, J_1(f_1{}^*)) = 0$ for all f in S. Thus $U = J_1{}^{-1}(S^{\perp})$. Let us now assume that

(a)$'_0$ *The maximal operator T of A in F is densely defined in F.*

Then we may define the adjoint operator T^* of T by

$$D(T^*) = \{f^*: \text{There exists } g^* \text{ in } F^* \text{ such that } (Tf, f^*) = (f, g^*) \text{ for all } f \text{ in } D(T)\},$$
$$T^*f^* = g^*.$$

In terms of T^*, we can formulate our first non-trivial assumption upon T, namely:

(a)$_1$ *The range of T^* is closed in F^*.*

If (a)$_1$ holds, it follows easily from the Hahn-Banach theorem that $S^{\perp} = (N(T))^{\perp} = R(T^*)$. Thus $U = J_1{}^{-1}(R(T^*))$, and for each $f_1{}^*$ in U, $J_1 f_1{}^* = T^*h^*$ for some h^* in $D(T^*)$.

We now formulate our second basic assumption:

(a)$_2$ $R(J_1) \cap R(T^*) = J_1(R(T_1{}^*))$.

Using these assumptions, we obtain the following simple proposition which we state in a general form without reference to maximal operators for use in a later section:

PROPOSITION 1. *Let G be an open subset of E^n, G_1 an open subset with compact closure in G, F a locally convex linear space of distributions on G with F and F^* as above, F_1 a locally convex linear space of distributions on G_1 such that the restriction map J of F into F_1 is continuous and has a dense image. Let J_1 be the injection mapping of $F_1{}^*$ into F^* given by $J_1 = J^*$.*

Suppose that T *is a densely defined linear mapping in* F, T_1 *is a densely defined linear mapping in* F_1 *such that* $J(N(T)) \subset N(T_1)$.

Suppose that:

(a)$_1$: *The range* $R(T^*)$ *is closed in* F^*.

(a)$_2$: $R(T^*) \cap R(J_1) = J_1(R(T_1^*))$.

Then $J(N(T))$ *is dense in* $N(T_1)$.

Proof of Proposition 1. As above, let $U = (J(T))^{\perp}$. We have already remarked that $U = J_1^{-1}(R(T^*))$. Since $J(N(T))$ is assumed to be contained in $N(T_1)$, in order to show that $J(N(T))$ is dense in $N(T_1)$, it suffices to show that $U \subset N(T_1)^{\perp}$. Suppose then that f_1^* is an element of U. Then $J_1 f_1^*$ lies in $R(J_1) \cap R(T^*)$ and $= J_1(T_1^* g_1^*)$ for some g_1^* in $D(T_1^*)$. Since the image of J is dense in F_1, $J^* = J_1$ is an injective map. Thus $f_1^* = T_1^* g_1^* \in R(T_1^*) \subset N(T_1)^{\perp}$, and the proof of Proposition 1 is complete.

Given Proposition 1, our problem is now the verification of the conditions (a)$'_0$, (a)$_1$, and (a)$_2$ under more concrete hypotheses.

Suppose again that, as before, T and T_1 are the maximal operators of a given linear differential operator A in F and F_1, respectively. Let us examine some conditions which suffice for the validity of (a)$'_0$, (a)$_1$, and (a)$_2$.

The case of (a)$'_0$ is the simplest. It suffices to assume the following:

(b)$_0$ *If* $R(G) = C^{\infty}{}_0(E^n)|_G$, $R(G)$ *is contained and dense in* F, *while multiplication by each coefficient* a_α *of* A *carries* F *into itself.*

(Indeed, since D^α for any α carries $R(G)$ into $R(G)$ which is contained in F, it follows that $R(G)$ is contained in the domain of T and $D(T)$ is thereby dense.)

To treat (a)$_1$, we consider first the case when G is precompact and obtain more general hypotheses by an iterative approximation. For some kinds of function spaces F on pre-compact sets G, there holds the following sufficient condition for the validity of (a)$_1$:

(b)$_1$ F^* *is a Frechet space, and there exists a linear subset* W *of* F^* *having a locally convex topology which turns* W *into a Frechet space such that the injection mapping of* W *into* F^* *is compact while* $D(T^*)$ *is contained in* W.

LEMMA 1. *Let* T^* *and* F^* *satisfy the condition* (b)$_1$. *Then the range of* T^* *is closed in* F^*.

Proof of Lemma 1. Let $G(T^*)$ be the graph of T^* in $F^* \times F^*$. Since

T^* is closed, $G(T^*)$ is a closed subspace of the Frechet spaces $F^* \times F^*$ and hence itself a Frechet space. Let P denote the mapping of $G(T^*)$ into F^* given by

$$P([f^*, T^*f^*]) = f^*.$$

Since $D(T^*)$ is contained in W, the range of P, which coincides with $D(T^*)$, lies in W. P is obviously a continuous linear mapping of $G(T^*)$ into F^*. Since the injection mapping of W into F^* being compact is a fortiori continuous, P is a closed mapping from $G(T^*)$ into W. On the other hand, $G(T^*)$ and W are both Frechet spaces and the closed graph theorem holds for mapping between Frechet spaces, so that P is a continuous mapping of $G(T^*)$ into W. Since the map P of $G(T^*)$ is the composition of the continuous map P of $G(T^*)$ into W with the compact injection map of W into F^*, it follows that P as a map of $G(T^*)$ into F^* is compact.

The conclusion of the Lemma now follows from Lemma 1 of Section 1 of [3] which declares that any mapping T^* for which the mapping of the graph upon the first factor is compact, has a closed range. Thus $R(T^*)$ is closed, and the proof of Lemma 1 is complete.

To apply Lemma 1, we must consider hypotheses of a more concrete character from which the criterion $(\mathrm{b})_1$ will follow. It is convenient for us therefore to introduce at this point various classes of function spaces F and F^* which we shall utilize in the application to approximation theorems for particular types of differential operators. We list below some of these spaces with the appropriate norms or semi-norms which define their topology:

(1) $C^r(G) = \{f \colon D^\alpha f \text{ exists and is continuous in } G \text{ for } |\alpha| \leqq r\}$,

$C^r(\bar{G}) = \{f \colon f \in C^r(G), \text{ For } |\alpha| \leqq r, D^\alpha f \text{ is uniformly bounded and continuous on } G\}$,

$$\| f \|_{C^r(G)} = \sum_{|\alpha| \leqq r} \sup_{x \in G} | D^\alpha f(x) |.$$

$C^r(\bar{G})$ is a Banach space with respect to the norm $\| f \|_{C^r(\bar{G})}$. $C^r(G)$ is a a Frechet space with respect to the topology defined by the semi-norms $\| f \|_{C^r(\bar{G}_1)}$ where G_1 runs through the open subsets having compact closure in G.

(2) Let $1 \leqq p < \infty$.

$L^p(G) = \{f \colon |f|^p \text{ is summable with respect to the Lebesque } n\text{-measure } dx \text{ on } G\}$,

$$\| f \|_{L^p} = \{ \int_G | f |^p \, dx \}^{1/p},$$

$L^\infty(G) = \{f : f$ is essentially bounded with respect to dx on $G\}$

$$\| f \|_{L^\infty(G)} = \text{ess. sup.}_{x \in G} | f(x) |,$$

$L^p{}_{\text{loc}}(G) = \{f : f \in L^p(G_1)$ for every open subset G_1 with compact closure in $G\}$.

$L^p(G)$ is a Banach space with respect to the norm $\| f \|_{L^p(G)}$.

$L^p{}_{\text{loc}}(G)$ is a Frechet space with respect to the topology defined by the family of semi-norms $\| f \|_{L^p(G_1)}$ for G_1 with compact closure in G.

$$(3) \quad W^{j,p}(G) = \{f : D^\alpha f \in L^p(G) \text{ for } |\alpha| \leqq j\}, \ (1 \leqq p < +\infty),$$

$$\| f \|_{j,p} = \{ \sum_{|\alpha| \leqq j} \| D^\alpha f \|^p{}_{L^p(G)} \}^{1/p},$$

$$W^{j,p}{}_{\text{loc}}(G) = \{f : f \in W^{j,p}(G_1) \text{ for } G_1 \text{ having compact closure in } G\}.$$

$W^{j,p}(G)$ is a Banach space with respect to $\| f \|_{j,p}$ while $W^{j,p}{}_{\text{loc}}(G)$ is a Frechet space with respect to the family of semi-norms $\| f \|_{W^{j,p}(G_1)}$.

(4) Let $0 < h < 1$.

$$C^{j,h}(G) = \{f : f \in C^j(\bar{G}), \frac{| D^\alpha f(x) - D^\alpha f(y) |}{| x - y |^h} \leqq C(f) \text{ for } |\alpha| \leqq j, \ x, y \in G\}.$$

$$\| f \|_{C^{j,h}(G)} = \| f \|_{C^r(\bar{G})} + \inf C(f),$$

$$C^{j,h}{}_{\text{loc}}(G) = \{f : f \in C^{j,h}(G_1) \text{ for every } G_1 \text{ with compact closure in } G\}.$$

$C^{j,h}(G)$ is a Banach space with respect to $\| f \|_{C^{j,h}(G)}$, while $C^{j,h}{}_{\text{loc}}(G)$ is a Frechet space with respect to the family of semi-norms $\| f \|_{C^{j,h}(G_1)}$, G_1 having compact closure in G.

We now remark the following consequences of the Sobolev Imbedding theorem (e.g. Lemma 5 of [2]):

LEMMA 2. *Let G be a precompact open subset of E^n for which there exists a finite open covering $\{N_k\}$ having the following property: For each k, there exists a Lipschitz continuous function $\xi_k(x)$ from $N_k \cap G$ to the unit sphere S in E^n and a positive constant h such that the cone*

$$C_k(x) = \{y : y = x + r\xi, 0 < r < h, | \xi - \xi_k(x) | < h, | \xi | = 1\}$$

is entirely contained in G for every x in $N_k \cap G$. (A condition of this type without prescription for the Lipschitz continuity of the axis is called a Sobolev cone condition.)

Then:

For $r^{-1} \geqq p^{-1} - (j-k)n^{-1}$, $0 \leqq k < j$, $1 < p < \infty$,

$$W^{j,p}(G) \subset W^{k,r}(G).$$

If $r^{-1} > p^{-1} - (j-k)n^{-1}$, *the imbedding map of* $W^{j,p}(G)$ *into* $W^{k,r}(G)$ *is compact.*

(ii) *For* $p^{-1} - (j-k)n^{-1} < 0$, $W^{j,p}(G) \subset C^k(\bar{G})$. *Moreover, if* $0 < h < 1$, *and*

$$p^{-1} < (j-k-h)n^{-1}$$

then $W^{j,p}(G) \subset C^{j,h}(G)$. *In both cases, the imbedding mapping is compact.*

(iii) *The imbedding map of* $W^{j,p}(G)$ *into* $W^{j-1,p}(G)$ *is compact. The imbedding map of* $C^{j,h}(G)$ *into* $C^{j-1,h}(G)$ *is compact.*

The conclusion of (iii) is obviously a specialization of (i) and (ii).

It follows from Lemma 2 in particular that if our space F^* is one of the spaces defined above, say in particular $L^r(G)$, then in order to verify that $(b)_1$ holds on a domain G which satisfies the condition of Lemma 2, it suffices to show that $D(T^*)$ lies in $W^{1,p}(G)$ for $p^{-1} < r^{-1} + n^{-1}$, or more generally that $D(T^*)$ lies in $W^{j,p}(G)$ for j and p such that $p^{-1} < r^{-1} + jn^{-1}$. In particular, we may take $p = r$, which always satisfies the inequalities, and we have the following: *If* $D(T^*) \subset W^{j,p}(G)$ *for a domain* G *satisfying the condition of Lemma 2,* $j > 0$, T^* *an operator in* $L^p(G)$ *with* G *precompact, then* T^* *has a closed range in* $L^p(G)$.

To pass from precompact domains G to general domains, we need an iterative approximation argument which we make precise in the following general formulation:

LEMMA 3. *Let* G *be an open set in* E^n, G_n *an increasing sequence of open sets exhausting* G *such that each* G_n *has compact closure in* G_{n+1}. *Let* F *be a Frechet space of functions on* G, F_n *a Banach space of functions on* G_n *such that the restriction map of* F_n *to the subdomain* G_{n-1}, *for each* n, *is a contraction operator carrying* F_n *into* F_{n-1}. *Suppose that* $F|_{G_n} = F_n$ *and that the topology on* F *is induced by the family of semi-norms* $\| f|_{G_n} \|_{F_n}$, *while* $f \in \mathfrak{D}'(G)$ *belongs to* F iff $f|_{G_n} \in F_n$ *for all* n.

Suppose that A *is a differential operator on* G *satisfying* $(b)_0$ *on* G *and on each of the* G_n, T *the maximal operator of* A *in* F, T_n *the maximal operator of* A *in* F_n *on* G_n. *Suppose that* $N(T_{n+1})|_{G_n}$ *is dense in* $N(T_n)$ *in the norm of* F_n *for each* $n \geqq 1$. *Then* $N(T)|_{G_1}$ *is dense in* $N(T_1)$ *in the norm of* F_1.

Proof of Lemma 3. Let f_1 be an element of $N(T_1)$, $\epsilon > 0$. We wish to show that there exist f in $N(T)$ with $\| f_1 - f|_{G_1} \| \leqq \epsilon$. We construct by induction starting with f_1 a sequence $\{f_n\}$ with $f_n \in N(T_n)$, $n \geqq 1$, such that $\| f_n - f_{n+1}|_{G_n} \| < \epsilon 2^{-n}$. This construction is possible since $N(T_{n+1})|_{G_n}$ is dense in $N(T_n)$ for $n \geqq 1$ by the hypothesis. From the iterative construction and the fact that the restriction mapping of each F_n on its predecessors is a contraction map, we find that for $j \geqq n$

$$\| f_{j+1}|_{G_n} - f_j|_{G_n} \|_{F_n} \leqq \| f_{j+1}|_{G_j} - f_j \|_{F_j} < \epsilon 2^{-j}.$$

Thus for $n \leqq j < k$, we have

$$\| f_j \|_{G_n} - f_k|_{G_n} \|_{F_n} \leqq \epsilon 2^{-(j-1)},$$

and hence $\{f_j|_{G_n}; j \geqq n\}$ is a Cauchy sequence in F_n for $n \geqq 1$. By the completeness of F_n, $f_j|_{G_n}$ converges to an element g_n of F_n, and we see at once that $g_{n+1}|_{G_n} = g_n$. The element f of F which we are looking for is obtained by setting $f|_{G_n} = g_n$. The distribution f thus obtained must lie in F since each g_n lies in F_n and is unambiguously defined since the values of g_n are concordant under restriction. Moreover since

$$\| f_j \|_{G_1} - f_1 \| \leqq \epsilon$$

for $j \geqq 1$, it follows by taking the limit on j that

$$\| g_1 - f_1 \|_{F_1} = \| f|_{G_1} - f_1 \| \leqq \epsilon.$$

Finally, each $f_j|_{G_n}$, $j > n$, satisfies the equation $Au = 0$ in G_n, and lies in $N(T_n)$. Hence so does their limit g_n. Thus $Af = 0$ in G_n for all n, i.e. $Af = 0$ in G, so that $f \in N(T)$ and the proof is complete.

Let us remark in connection with the application of Lemma 3 that the conditions of the first paragraph of the hypothesis of Lemma 3 concerning the spaces F and the F_n will be satisfied for any of the following:

(1) $F = C^r(G)$, $F_n = C^r(G_n)$.

(2) $F = L^p{}_{\mathrm{loc}}(G)$, $F_n = L^p(G_n)$.

(3) $F = W^{j,p}{}_{\mathrm{loc}}(G)$, $F_n = W^{j,p}(G_n)$.

(4) $F = C^{j,h}{}_{\mathrm{loc}}(G)$, $F_n = C^{j,h}(G_n)$.

The crucial topological hypothesis that we shall utilize upon the relation between G and its subset G_1 is the property that $G - G_1$ has no compact components. The following lemma, therefore, is necessary as an auxiliary tool for Lemma 3 to reduce our approximation arguments to the case of precompact sets G.

LEMMA 4. *Let G be an open subset of E^n, G_1 an open subset with compact closure in G such that $G - G_1$ has no compact components. Then there exists an increasing sequence $\{G_k\}$ of open subsets of G starting with G_1 such that G is the union of the G_k, each G_k has compact closure in G_{k+1}, and for each k, $G_{k+1} - G_k$ has no compact components.*

Proof of Lemma 4. Let $\{G'_k\}$ be an arbitrary increasing sequence of open subsets of G which together exhuast G and such that G'_k has compact closure in G'_{k+1}. We shall construct the sequence $\{G_k\}$ by induction so that for each k, G_k contains G'_k, G_k has compact closure in G_{k+1}, $G_{k+1} - G_k$ has no compact components, and in addition $G = G_k$ has no compact components. Suppose that $\{G_1, \cdots, G_k\}$ have already been constructed, and let us consider the problem of continuing the construction to obtain G_{k+1}.

Let V be a smoothly bounded set with compact closure which contains G'_{k+1} and the closure of G_k. (By smoothly bounded, we mean more precisely that V is bounded by a finuite number of polytopes.) To obtain G_{k+1}, we need only choose a set V_1 which has compact closure in G, which contains V, and such that $V_1 - G_k$ and $G - V_1$ have no compact components. For such a V_1, we take the union of V with all the compact components of $G - V$. Since each component of $G - V_1$ must be contained in a component of $G - V$ and indeed must equal one of the non-compact components of $G - V$, it follows that $G - V_1$ has no compact components. The closure of V_1 is certainly compact, being the union of a finite family of compact sets. We need only show then that $V_1 - G_k$ has no compact components and that V_1 is open.

Let us show first that V_1 is open. Since V is smoothly bounded (in the special sense mentioned above), each point of the boundary of V has a neighborhood which intersects only a finite number of components of $G - V$. Since the boundary of each compact component of $G - V$ is contained in the boundary of V, it follows that there are only a finite number of compact components of $G - V$ and that they have a positive distance from the union of the other components of $G - V$. Hence V_1 is open.

To show that $V_1 - G_k$ has no compact components, let C be any compact component of $V_1 - G_k$. Since G_k has compact closure in V_1, we may choose a neighborhood V_2 of $C \cup G_k$ such that $\bar{V}_2$ is contained in V_1. C is a component of $\bar{V}_2 - G_k$, which is a compact set, and hence C is the intersection of a sequence of open and closed sets of $\bar{V}_2 - G_k$. We assert that C is a component of $G - G_k$. Let $\{U_s\}$ be a sequence of open and closed sets of $\bar{V}_2 - G_k$ whose intersection equals C. Since each U_s is closed in the compact set $\bar{V}_2 - G_k$, it is closed in $G - G_k$. On the other hand, $U_s \cap \operatorname{bdry}(V_2)$ is a family of compact sets having null intersection, so that we may find a finite intersection of sets U_s which is contained in $V_2 — G$. Since we may then assume, replacing U_s by their intersections with this finite family, that each U_s is contained in $V_2 — G_k$ and being open in $\bar{V}_2 — G_k$, it must be open in the smaller set $V_2 — G_k$. But $V_2 — G_k$ is open in $G — G_k$, so that U_s is open in $G — G_k$. Hence C is the intersection of a family of sets U_s open and closed in $G — G_k$ and hence C is a compact component of $G — G_s$. There are no such components by the inductive hypothesis, and the proof of Lemma 4 is complete.

Section 2. As we have remarked in Proposition 1, the proof of the density of $N(T)$ in $N(T_1)$ on restriction to the open subset G_1 follows from the two conditions:

(a)$_1$. *$R(T^*)$ is closed in F^*.*

(a)$_2$ $R(T^*) \cap R(J_1) = J_1 R(T_1^*))$.

A deeper investigation of hypotheses under which these conditions hold demands therefore a more detailed consideration of the structure of the operators T^* and T_1^*.

DEFINITION. *Let T'_{00} be the operator in F^* with $D(T'_{00}) = C^\infty{}_o(G)$, $T'_{00}\psi = A'\psi$ for $\psi \in D(T'_{00})$.*

Because of our assumption that $A'\psi$ lies in F^* for each ψ in $C^\infty{}_o(G)$, the operator T'_{00} is well-defined.

LEMMA 5. *T^* is the closure of the operator T'_{00} as an operator with domain and range in F^*.*

Proof of Lemma 5. By a familiar argument on annihilators in linear spaces it suffices to show that for a pair of elements f and g of F such that

$$(f, T'_{00}\psi) = (g, \psi), \text{ all } \psi \text{ in } C^\infty{}_c(G)$$

f will lie in $D(T)$ and $Tf = g$, and conversely. But this is precisely the definition of the maximal operator T. Q.E.D.

The operator T^* thus identified with the realization of the differential operator A' acting on a 'minimal' family of functions satisfying the largest possible family of boundary conditions in G is called the *minimal operator* of A' in F^*.

LEMMA 6. *$J_1(D(T_1^*)) \subset D(T^*)$. For f_1^* in $D(T_1^*)$,*

$$T^*(J_1^*) = J_1(T_1^* f_1^*).$$

Proof of Lemma 6. Consider the graphs $G(T_1{}^*)$ and $G(T^*)$. The mapping $J_1 \times J_1$ carries the dense subset $G(T'_{00,G_1})$ of $G(T_1{}^*)$ into $G(T'_{00}) \subset G(T^*)$, where T'_{00,G_1} is the operator analogous to T'_{00} defined for $F^*{}_1$ and G_1 rather than F^* and G. $J_1 \times J_1$ is continuous. Hence it carries $G(T_1{}^*)$, the closure of $G(T'_{00,G_1})$, into $G(T^*)$. The assertions of Lemma 6 are precisely equivalent to the last aseertion. Q. E. D.

PROPOSITION 2. *Let G_1, G, and G_2 be three open subsets of E^n with G_1 having compact closure in G, G' having compact closure in G_2. Let F_1, F and F_2 be three spaces on G_1, G, and G_2, respectively, with the pairs (F_1, F) and (F, F_2) satisfying the conditions of our preceding discussion, $F_1{}^*$, F^*, and $F_2{}^*$ their dual spaces. Let A be a linear differential operator on G_2, A' its formal adjoint, with A satisfying* $(\mathrm{b})_0$ *on G_2. Suppose that* $(\mathrm{b})_1$ *holds on G and that the following conditions are satisfied:*

$(\mathrm{c})_1$. *Every h^* in $D(T_2{}^*)$ which has its support in the closure of G and which satisfies the equation $A'h^* = 0$ on $G_2 - G_1$ must have its support in the closure of G_1.*

$(\mathrm{c})_2$. *Every h^* in $D(T_2{}^*)$ which has its support in the closure of G_1 lies in $J_2(D(T_1{}^*))$, where J_2 is the natural injection of $F_1{}^*$ into $F_2{}^*$.*

Then if T is the maximal operator of A in F on G, T_1 the maximal operator of A in F_1 on G_1, we have $N(T)\,|_{G_1}$ dense in $N(T_1)$ in the F_1-topology.

Proof of Proposition 2. Since $(\mathrm{b})_1$ holds on G, it suffices by Proposition 1 to verify that $R(T^*) \cap R(J_1) = J_1(R(T_1{}^*))$. By Lemma 6, $J_1(R(T_1{}^*)) \subset R(T^*) \cap R(J_1)$. Hence it suffices to prove the opposite inclusion.

Suppose then that $f^* \in R(T^*) \cap R(J_1)$. Then f^* has its support in the closure of G_1, and there exists g^* in $D(T^*)$ such that $T^*g^* = f^*$. Let h^* be the injection of g^* into $F_2{}^*$. Then by Lemma 6 for the pair of domains G and G_2, $T_2{}^*h^*$ has its support in $\bar{G}_1$ and h^* has its support in $\bar{G}$. By $(\mathrm{c})_1$, h^* must have its support in $\bar{G}$. By $(\mathrm{c})_2$, $h^* = J_2(k^*)$ with $k^* \in D(T_1{}^*)$. Thus g^* and $J_1(k^*)$ have the same injection into $F_2{}^*$, so that $g^* = J_1(k^*)$. Hence $f^* = T^*g^* = T^*(J_1k^*) = J_1(T_1{}^*k^*)$ must lie in $J_1(R(T_1{}^*))$ and the proof of Proposition 2 is complete.

Let us now reformulate Proposition 2 slightly to make it more directly applicable to approximation problems for solutions of differential equations.

THEOREM 1. *Let G_1, G, and G_2 be three open sets in E^n, $\bar{G}_1$ compact in G, $\bar{G}$ compact in G_2. Let F_1, F, and F_2 be three locally convex linear spaces on G_1, G, and G_2, respectively, as above with dual spaces $F_1{}^*$, F^*, and*

$F_2{}^*$. *Suppose that the restriction map is a continuous linear map of* F *on a dense subset of* F_1 *and of* F_2 *on a dense subset of* F. *Let* A *be a linear partial differential operator of order* r *on* G_2 *with formal adjoint* A' *such that* A' *has coefficients in* $F_2{}^*$, $R(G_2)$ *is dense in* F_2, *and multiplication by the coefficients of* A *carry* $R(G_2)$ *into* F_2. *Suppose further that*:

(i) *The domain of the minimal operator of* A' *in* F^* *on* G *lies in a linear subset* W *of* F^* *which is a Banach space with respect to a norm such that the injection mapping of* W *into* F^* *is compact.*

(ii) *If* u *is a distribution in* $F^*{}_2$ *which lies in the domain of the minimal operator of* A' *in* $F^*{}_2$ *and if* $A'u = 0$ *on* $G_2 - \bar{G}_1$ *then* $u \in C^r(G_2 - \bar{G}_1)$.

(iii) *If* $u \in C^r(G_2 - \bar{G}_1)$ *with* $A'u = 0$ *on* $G_2 - \bar{G}_1$ *and if* $u = 0$ *of* $G_2 - \bar{G}$, *then* u *must be identically zero* in $G_2 - \bar{G}_1$.

(iv) *Every* u *in* $F^*{}_2$ *which has its support in* $\bar{G}_1$ *and is in the domain of the minimal operator of* A' *in* F_{02} *is the injection into* G_2 *of an element of the domain of the minimal operator of* A' *in* G_1.

Then the solutions S *in* F *of the equation* $Au = 0$ *in* G *upon restriction to the open subset* G_1 *are dense in the* F_1*-topology in the solutions* S_1 *in* F_1 *of the equation* $Au_1 = 0$ *in* G_1.

Proof of Theorem 1. (ii) and (iii) together obviously imply $(c)_1$, (i) is equivalent to $(b)_1$ (iv) is equivalent to $(c)_2$, so that Theorem 1 follows immediately from Proposition 2.

Section 3. We turn now to the principal topic of the present paper, the proof of an approximation theorem in uniform norm of the Walsh-Mergelyan type for solutions of a general linear elliptic equation.

Let A be a linear elliptic operator of order r on the domain G of E^r of the form

$$A = \sum_{|\alpha| \leqq r} a_\alpha(x) D^\alpha$$

with coefficients a_α lying in $C^1(G)$. Let A' be the formal adjoint of A, $A' = \sum_{|\alpha| \leqq r} a'_\alpha D^\alpha$, with coefficients a'_α which again lie in $C^1(G)$. We shall assume that A' satisfies the condition of uniqueness for the Cauchy problem in the small on G, i. e.

$(\mathrm{U})_s$. *If* $u \in C^r(G')$ *for a connected open subset* G' *of* G *with* $A'u = 0$ *on* G' *and if* u *vanishes on a non-vacuous open subset of* G', u *must vanish identically on* G'.

We remark that $(\mathrm{U})_s$ is satisfied by operators A' with analytic coefficients and, by a theorem of Calderon [6], $(\mathrm{U})_s$ holds for operator A' with real coefficients having distinct complex characteristics.

We consider the specialization of the approximation problem of Sections 1 and 2, where the space F is $C^0(G)$ and the space F_1, on the open subset G_1 on which the approximation is to take place, is $C^0(\bar{G}_1)$, the Banach space of continuous functions on the closure of G_1. As is well known from the representation theorem of F. Riesz and its variants, the dual spaces of these two linear spaces are $F^* = M_c(G)$ and $F^*_1 = M(\bar{G}_1)$ where $M_c(G)$ is the space of finite Radon measures on G with compact support in G while $M(\bar{G}_1)$ is the space of finite Radon measure on $\bar{G}_1$. In order to deal with the minimal operators of A' in these two spaces, we remark that by the Sobolev Imbedding Theorem in its local form, every distribution lying in $W^{1,p}{}_{\mathrm{loc}}(G)$ must lie also in $C^0(G)$ provided that $p > n$, and the imbedding mapping of $W^{1,p}{}_{\mathrm{loc}}(G)$ into $C^0(G)$ is a continuous linear map (as follows also by the closed graph theorem from the first assertion). In addition, the image of the imbedding mapping is dense.

Let us introduce a representation of the dual space $(W^{1,p}{}_{\mathrm{loc}}(G))^*$ by distributions having compact support in G with the distribution pairing going over into the pairing between the dual spaces. Thus if $q = p(p-1)^{-1}$, we set

$$W^{-1,q}{}_c(G) = \{u : u \in \mathcal{E}'(G), |(\psi, u)| \leqq C_u \| \psi \|_{W^{1,p}(G)} \text{ for all } \psi \text{ in } C^\infty(G)\}.$$

$$\| u \|_{W^{1,q}(G)} = \inf C_u.$$

Then from the results of the preceding paragraph, it follows that $M_c(G) = (C^0(G))^*$ is contained in $W^{-1,q}{}_c(G)$ and the imbedding mapping of $M_c(G)$ into $W^{-1,q}{}_c(G)$ is bounded under the assumption $1 < q < n(n-1)^{-1}$. The minimal operator T^* of A' in $M_c(G)$, which is the closure in $M_c(G)$ of the operator A' restricted to $C^\infty{}_c(G)$, must therefore be contained in the minimal operator of A' in $W^{-1,q}{}_c(G)$. By the results of Section 2 of [3] specifically Theorem (2.7)), the domain of the minimal operator of A' in $W^{1,q}{}_c(G)$ is contained in $W^{r-1,q}{}_c(G)$. We have therefore proved:

Lemma 7. *Let A be a linear elliptic operator of order r such that A and A' have coefficients in $C^1(G)$. Then the domain of the minimal operator T^* of A' in $M_c(G)$, the space of Radon measures with compact support in G, is contained in $W^{r-1,q}{}_c(G)$ for any q with $1 < q < n(n-1)^{-1}$.*

To obtain our uniform approximation theorem, we must impose a regularity condition upon the domain G_1 on which the approximation is to take place.

DEFINITION 2. *Let G_1 be an open subset of E^n. Then G_1 is said to be firmly regular if there exists an open covering $\{N_k\}$ of the boundary of G_1 such that for each k, there exists a positive constant h_k and ξ_k in E^n with $|\xi_k| = 1$, for which the punctured cone*

$$C'_x = \{y : y = x + r\xi, 0 < r < h, |\xi - \xi_x| < h, |\xi| = 1\}$$

lies in G for any point x of $N_k \cap \bar{G}$.

For a precompact G_1, the definition of firm regularity is obviously a strengthening of the Sobolev cone condition. In particular, for any precompact, firmly regular G_1, $W^{j,q}(G_1)$ is compact in $L^q(G_1)$ for any $j > 0$, $q > 1$. We have moreover the following important property of firmly regular open sets:

LEMMA 8. *Let G be an open subset of E^n, G_1 a firmly regular open set with compact closure in G, A' a differential operator of order r with coefficients in $C^1(G)$. Then:*

(a) *Let $u \in W^{r-1,q}(G)$ with support in $\bar{G}_1$ such that $A'u \in M(\bar{G}_1) \subset M_o(G)$. Then there exists a sequence $\{\psi_s\}$ in $C^\infty{}_o(G_1)$ such that $\psi_s \to u$ in $W^{r-1,q}(G_1)$, $A'(\psi_s) \to A'(u)$ in the weak* topology on $M(\bar{G}_1)$.*

(b) *Let $u \in W^{r-1,q}(G)$ with support in $\bar{G}_1$ such that $A'u \in L^q(G_1) \subset L^q(G)$. Then there exists a sequence $\{\psi_s\}$ in $C^\infty{}_o(G_1)$ such that $\psi_s \to u$ in $W^{r-1,q}(G_1)$, $A'(\psi_s) \to A'u$ in $L^q(G_1)$.*

Proof of Lemma 8. We first form an infinitely differentiable partition of unity $\{\eta_k\}$ corresponding to the covering of $\bar{G}_1$ by the family $\{N_k\}$ together with a single interior open set in G_1. The function $u = \sum_k u_k$, where $u_k = \eta_k u$. To approximate u in the sense of Lemma 8, it suffices to approximate each of the functions u_k. Each u_k lies in $W^{r-1,q}(G_1)$ and $A'(u_k) = \eta_k u + v_k$, where v_k lies in $L^q(G_1)$. Hence $A'u_k$ has the same properties as $A'u$ in both (a) and (b). Thus, replacing u by u_k, we may assume that u either has compact support in the interior of G_1 or else in one of the neighborhoods N_k of the definition of firm regularity. We shall consider only the latter case explicitly, since the former can be dealt with by a simpler version of the same argument.

Suppose then that u has its support in $N_k \cap \bar{G}_1$. By Definition 2, there exists a constant $h_k > 0$ and a unit vector ξ_k such that all points y of the form

$$y = x + r\xi, \quad 0 < r < h, \quad |\xi - \xi_k| < h, \quad |\xi| = 1,$$

lie in G_1 for x in $N_k \cap \bar{G}_1$. We form a function j, depending implicitly upon k, such that $j \in C^\infty{}_o(E^n)$, $j(x)$ has its support in the set

$$\{x : 1 \leqq |x| \leqq 2, \ |\xi_k - x/|x|| < h\}, \text{ and } \int j(x)\,dx = 1.$$

For $\epsilon > 0$, set $j_\epsilon(x) = \epsilon^{-n} j(\epsilon^{-1}x)$. From the properties of j, it follows that j_ϵ also lies in $C^\infty{}_c(E^n)$, $\int j_\epsilon(x)\,dx = 1$, while the support of j_ϵ is contained in the set

$$C_\epsilon = \{x \colon \epsilon \leqq |x| \leqq 2\epsilon, |\epsilon_k - x/|x|| < h\}.$$

Consider

$$v_\epsilon(x) = \int j_\epsilon(x-y) u_k(y)\,dy, \ (0 < \epsilon < h).$$

The support of v_ϵ lies in $C_\epsilon + N_k \cap \bar{G}_1$, which for each positive ϵ is a compact subset of G_1. The function v_ϵ is infinitely differentiable and hence lies in $C^\infty{}_c(G_1)$. To complete the argument, it suffices to show that $v_\epsilon \to u_k$, $A'v_\epsilon \to A'u_k$ in the senses sought in parts (a) and (b) of Lemma 7.

Since u_k lies in $W^{r-1,q}(G)$ and has its support in $\bar{G}_1$, it follows by a standard argument that v_ϵ will converge to u_k in $W^{r-1,q}(G_1)$. The principal point to be dealt with is the convergence of $A'v_\epsilon$. For this purpose, we let J_ϵ denote the convolution operator with kernel j_ϵ, and consider

$$r_\epsilon = A'v_\epsilon - J_\epsilon(A\ u_k) = A'J_\epsilon u_k - J_\epsilon A'u_k.$$

Since by standard arguments, $J_\epsilon(A'u_k)$ will converge to $A'u_k$ in the senses sought in (a) and (b) as $\epsilon \to 0$, in order to complete the proof of the Lemma, it suffices to show that $r_\epsilon \to 0$ in $L^q(G_1)$.

The function r_ϵ is the sum of a finite number of terms of the form

$$a(x) D^\alpha J_\epsilon u_k - J_\epsilon(a(x) D^\alpha u_k), \ |\alpha| \leqq r,$$

with $a(x) \in C^1(\bar{G}_1)$. Since the differentiation operators commute with J_ϵ in thois case we may absorb $(r-1)$ of the derivatives into the function u_k, and using the assumption that u_k lies in $W^{r-1,q}(G_1)$, it suffices to consider terms of the form

$$s_\epsilon(v) = a(x) D_k J_\epsilon(a(x) D_k v)$$

for v in $L^q(G_1)$ and show that s_ϵ converges to zero in $L^q(G_1)$. However, this follows from a well-known argument of Friedrichs [7] and the proof of the Lemma is complete.

A simple variant of the proof of Lemma 4 yields:

LEMMA 4′. *Let G be an open subset of E^n, G_1 a firmly regular open subset with compact closure in G such that $G - G_1$ has no compact components. Then there exists an increasing sequence $\{G_k\}$ of open subsets of G beginning with G_1 such that each G_k is firmly regular, G is the union of the G_k, each G_k*

has compact closure in G_{k+1}, *and for each* k, $G_{k+2}-G_k$ *has no compact components.*

Using Lemmas 7 and 8 together with Lemma 4′, we obtain:

THEOREM 2. *Let* G *be an open subset of* E^n, G_1 *a firmly regular open subset with compact closure in* G. *Suppose that* A *is a linear elliptic differential operator on* G *with coefficients in* $C^1(G)$ *such that its formal adjoint* A' *has its coefficients in* $C^1(G)$. *Suppose that* $G-G_1$ *has no compact components and that* A' *satisfies the condition* $(\mathrm{U})_s$ *for uniqueness in the Cauchy problem in the small on* G. *Let* S *be the family of solutions* u *in* $C^r(G)$ *of the equation* $Au=0$ *in* G, S_1 *the family of functions* u_1 *in* $C^0(\bar{G}_1)\cap C^r(G_1)$ *which satisfy the equation* $Au_1=0$ *in* G_1. *Then the restrictions of the functions of* S *to* G_1 *are dense in* S_1 *in the norm of* $C^0(\bar{G}_1)$, *i.e. given a solution* u_1 *in* $C^0(\bar{G}_1)\cap C^r(G_1)$ *of the equation* $Au_1=0$ *in* G_1, *for every* $\epsilon>0$, *there exists* u *in* $C^r(G)$ *with* $Au=0$ *in* G *such that*

$$\sup_{x\in G_1}|u(x)-u_1(x)|<\epsilon.$$

Proof of Theorem 2. By Lemma 4, there exists a sequence of firmly regular open sets $\{G_k\}$ exhausting G, starting with G_1, and such that $G_{k+1}-G_k$ has no compact components and $\bar{G}_k$ is a compact subset of G_{k+1}. By Lemma 3, it suffices to show the uniform approximability of u_k in $C^0(\bar{G}_k)$ with $Au_k=0$ in G_k by u_{k+1} in $C^0(\bar{G}_{k+1})$ with $Au_{k+1}=0$ on G_{k+1}. It suffices as well to approximate instead by u_{k+2} in $C^0(G_{k+2})$ with $Au_{k+2}=0$ on G_{k+2}. Thus we may restrict ourself to a situation of the type considered in Proposition 2, where we have three open sets G_1, G, and G_2 with G_1 and G firmly regular, G_1 having compact closure in G, G having compact closure in G_2, G_2-G having no compact components, and show that if A is a linear elliptic differential operator on G_2 with A and A' having coefficients in $C^1(G_2)$ and A' satisfying $(\mathrm{U})_s$, then the solutions u of $Au=0$ in $C^0(G_2)$ are dense in $C^0(\bar{G}_1)$. (The fact that under our regularity hypotheses on the coefficients of A and A', the solutions of $Au=0$ in $C^0(G_2)$ lie in $C^r(G_2)$ follows from Theorem (2.7) of [3].)

To obtain this last result, we apply Theorem 1, and need therefore to verify the hypotheses of Theorem 1 for $F_2=C^0(G_2)$, $F=C^0(\bar{G})$, $F_1=C^0(\bar{G}_1)$, $F_2^*=M_c(G_2)$, $F^*=M(\bar{G})$, $F_1^*=M(\bar{G}_1)$. By Lemma 7, the domain of the minimal operator T^* of A' in F^* lies in $W^{r-1,q}{}_c(G_2)\subset W^{r-1,q}(G)$, which by the firm regularity of G is a compact subset of $M(\bar{G})$. Thus assumption (i) holds. Since A' is elliptic and both A and A' have coefficients in $C^1(G_2)$, it follows from Theorem (2.7) of [3] that every solution u of $A'u=0$ in

$G_2 - G_1$ lying in $W^{-1,q}(G_2)$ must lie in $C^r(G_2)$. Therefore, assumption (ii) holds. Assumption (iii) is seen by a standard argument to be equivalent to $(\mathrm{U})_s$ under the assumption that $G_2 - G_1$ has no compact components. The only assumption which remains for verification is thereby assumption (iv).

To verify (iv), let us suppose that u lies in $M_c(G_2)$ has its support in $\bar{G}_1$, and lies in the domain of the minimal operator of A' in $M_c(G_2)$. By Lemma 7, u must lie in $W^{r-1,q}{}_c(G_2)$ for $q < n(n-1)^{-1}$. Applying Lemma 8, we know therefore that there exists a sequence $\{\psi_s\}$ in $C^\infty{}_c(G_1)$ such that $\psi_s \rightarrow u$ in $L^q(G_1)$, $A'\psi_s \rightarrow A'u$ in the weak* topology in $M(\bar{G}_1)$. Hence the element $[u, A'u]$ lies in the weak* closure of the graph of the minimal operator $T_1{}^*$ of A' in $F_1{}^*$. However, $T_1{}^*$ being an adjoint operator, has a graph which is weak* closed. Thus $u \in D(T_1{}^*)$, (iv) is proved, and the proof of Theorem 2 is complete.

Section 4. In connection with Theorem 2, S. Agmon has conjectured in a conversation with the writer that a related result which is a slight variant of Theorem 2 should hold under stronger hypotheses. We establish this result below in Theorem 3, and give a simpler variant of the preceding proof under the stronger hypotheses of Theorem 3.

Let A be a linear elliptic differential operator on an open set G of E^n with coefficients in $C^\infty(E^n)$. Then if *both* A and A' satisfy the condition $(\mathrm{U})_s$ of uniqueness in the Cauchy problem in the small, Malgrange has shown in the last section of [11] that there exists a bi-regular elementary solution $e(x,y)$ for A on G, i.e. such that $e(x,y)$ is in C^∞ in both variables off the diagonal and satisfies the equations $A_x e(x,y) = \delta_y$, $A'_y \bar{e}(x,y) = \delta_x$. (A corresponding result under weaker regularity conditions on the coefficients has been obtained by the writer in Section 3 of [3], but we shall not refer to the latter result to avoid a complicated statement of the hypotheses).

THEOREM 3. *Let A be a linear elliptic differential operator on an open subset G of E^n with coefficients in $C^\infty(E^n)$ such that both A and A' satisfy the condition $(\mathrm{U})_s$ of uniqueness in the Cauchy problem in the large, $e(x,y)$ any bi-regular fundamental solution for A on G in the sense given above. Let G_1 be a firmly regular open subset with compact closure in G, G_3 an open subset with compact closure in G such that $\bar{G}_3$ and $\bar{G}_1$ do not intersect. Suppose that $G - G_1$ has no compact components. Then every solution u_1 of the equation $Au_1 = 0$ on G_1 with u_1 lying in $C^\infty(G_1) \cap C^0(\bar{G}_1)$ can be uniformly approximated on G_1 by a finite linear combination of functions of the form $e(x, y_k)$ with $y_k \in \bar{G}_3$.*

Proof of Theorem 3. Consider the closure in $C^0(\bar{G}_1)$ of the finite linear combinations of functions $e(x, y_k)$, $y_k \in \bar{G}_3$. All these functions when restricted to G_1 lie in the space S_1 of solutions of the equation $Au_1 = 0$ with $u_1 \in C^\infty(G_1) \cap C^0(\bar{G}_1)$, and S_1 is a closed subspace of $C^0(\bar{G}_1)$. To show the density of these linear combinations, it suffices to show that any measure m in $M(\bar{G}_1)$ which annihilates all these linear combinations must also annihilate S_1.

Let f be any function in $C^0(\bar{G}_3)$. The function $v(x)$ on $\bar{G}_1$ given by

$$v(x) = \int_{\bar{G}_3} e(x, y) f(y)\, dy$$

can be uniformly approximated on G_1 by a finite linear combination of $e(x, y_k)$, $y_k \in \bar{G}$. Hence

$$\int_{G_1} v(x)\, dm(x) = 0.$$

Using Fubini's theorem, this last integral can be written

$$\int f(y) \{ \int e(x, y)\, dm(x) \} dy = 0.$$

Thus the continuous function

$$w(y) = \int e(x, y)\, dm(x)$$

on $\bar{G}_3$ annihilates all continuous f and must vanish identically on G_3. But $w(y)$ is a solution on $G - \bar{G}_1$ of the equation $A'w = 0$, and its vanishing on the non-trivial open subset G_3 of $G - \bar{G}_1$, the uniqueness in the Cauchy problem in the small for A', and the fact that $G - G_1$ has no compact components together imply that $w(y)$ vanishes identically on $G - \bar{G}_1$. Hence w has its support in G_1.

On all of G, however, w is a distribution solution of the equation $A'w = m$. By Lemma 7, w must lie in $W^{r-1,q}(G_1)$ for $q < n(n-1)^{-1}$. By Lemma 8, there exists a sequence $\{\psi_s\}$ in $C^\infty{}_0(G_1)$ such that $\psi_s \to w$ in $W^{r-1,q}(G_1)$, $A'\psi_s \to m$ in the weak* topology on $M(\bar{G}_1)$. Let u_1 be any function in S_1. Then

$$\int_{G_1} u_1(x)\, dm(x) = \lim_s \int_{G_1} u_1(x) A'\psi_s(x)\, dx = \lim_s \int_{G_1} (Au)(x)\psi_s(x)\, dx = 0,$$

and m annihilates S_1. Thus the proof of Theorem 3 is complete.

Section 5. We may generalize Theorem 2 in a different direction by considering approximating functions u which are not only solutions of $Au = 0$ on G but satisfy homogeneous regular elliptic boundary conditions on a proper part of the boundary of G. The discussion of the most general results of this type is given in a forthcoming paper of the writer ([15]) which studies the analytic properties of solutions of general elliptic boundary value problems in great detail. We shall content ourselves here with the simplest case, that of null Dirichlet boundary conidtions on a proper part of the boundary. The discussion of this case suffices to indicate the general principles of the argument.

THEOREM 4. *Let G be a bounded open subset of E^n contained as a proper part of another bounded open set G'. Suppose that the boundary Γ of G is a manifold of class C^{2m} in E^n, and let G_1 be a firmly regular open subset with compact closure in G such that $G' - G_1$ has no compact components, Γ_0 a proper open subset of which is bounded by a finite number of $(n-2)$-dimensional manifolds of class C^{2m} in Γ and such that Γ_0 lies on the boundary of G'. Let A be a uniformly elliptic, regularly elliptic differential operator of order $2m$ on G' with coefficients in $C^1(\bar{G}')$ and such that A' has coefficients in $C^1(\bar{G}')$. Suppose that A' satisfies the condition $(\mathrm{U})_s$ for uniqueness in the Cauchy problem in the small on G'.*

Let S be the family of solutions u in $C^{2m}(G) \cap C^{m-1}(G \cup \Gamma_0)$ of the equation $Au = 0$ in G with null Dirichlet data of order m on Γ_0, S_1 the family of solutions u_1 in $C^{2m}(G_1) \cap C^0(\bar{G}_1)$ of $Au_1 = 0$ in G_1. Then every u_1 in S_1 is uniformly approximable on G_1 by functions u in S.

Proof of Theorem 4. Since the line of argument parallels the preceding proofs of Theorems 1 and 2 and is given in a more general context in [5], we shall indicate the proof of Theorem 4 in a very brief form. We begin with the remark that S is the nullspace of the operator T in $W^{1,p}(G)$, $(p > n)$, whose domain consists of all v in $W^{2m+1,p}{}_{\mathrm{loc}}(G \cup \Gamma_0)$ with null Dirichlet data on Γ_0 and such that Au lies in $W^{1,p}(G)$. The adjoint T^* of T, by the results of [2], is the operator in $W^{-1,q}(G)$ whose domain consists of all v in $W^{2m-1,q}(G)$, $q = p(p-1)^{-1}$, which satisfy null Dirichlet boundary conditions on Γ_0 and null Cauchy boundary conditions on $\Gamma - \Gamma_0$. Furthermore, because the domain of T^* lies in $W^{2m-1,q}(G)$ and G is a Sobolev domain, it follows from Lemma 1 that the range of T^* is closed in $W^{-1,q}(G)$.

Let m be a Radon measure on $\bar{G}_1$ which annihilates S. Considering m as an element of $M_0(G)$, m lies in $W^{-1,q}(G)$ and annihilates the nullspace of T. Since T^* has a closed range, m must lie in the range of T^*. Hence there exists v in $W^{2m-1,q}(G)$ which satisfies null Cauchy boundary conditions on $\Gamma - \Gamma_0$ such that $A'v = m$. Extending v to be a null on $G - G'$, we see that v is an element of $W^{2m-1,q}(G')$ which vanishes on $G' - \bar{G}$ and is a distribution solution of $A'v = 0$ on $G' - \bar{G}_1$. By the regularity theory of elliptic operators, (Theorem (2.7) of [3]), v must lie in $C^{2m}(G'-\bar{G}_1)$. On the other hand, by the condition $(\mathrm{U})_s$ of uniqueness in the Cauchy problem in the small for A' on G', it follows that v must be null outside of $\bar{G}_1$. Hence v is an element of $W^{2m-1,q}(G)$ with $A'v \in M(\bar{G}_1)$ and such that v has its support in $\bar{G}_1$. Since G_1 is firmly regular, it follows from Lemma 8 that there exists a sequence $\{\psi_s\}$ in $C^\infty{}_c(G_1)$ such that $A'\psi_s \rightarrow A'v = m$ in the weak* topology in $M(\bar{G}_1)$. Thus for all u_1 in S_1,

$$\int_{\bar{G}_1} u_1(x)dm(x) = \lim_s \int_{G_1} u_1(x)(A'\psi_s)(x)dx = \lim_s \int_{G_1} (Au_1)(x)\psi_s(x)dx = 0,$$

and the proof of Theorem 4 is complete.

Section 6. We turn in the present section to the treatment of approximation theorems in the spaces $W^{j,p}(G_1)$. For elliptic operators, we have already given a discussion of results of this type in [3], but we shall strengthen the discussion given there to such an extent as to be able to treat general classes of non-elliptic operators, specifically hyperbolic operators and principal normal operators having pseudo-convex functions in the sense of Hörmander [8]. Let us remark that in this passage to non-elliptic operators we no longer obtain uniform approximation theorems for solutions of $Au = 0$ for u lying in $C^0(G_1)$, but by the Sobolev Imbedding theorem, we do have uniform approximation for solutions u_1 of $Au_1 = 0$ for u_1 lying in $W^{j,2}(G_1)$ with $j > n/2$.

Let G be a firmly regular open set in E^n. Let j be a non-negative integer, p a real number with $1 < p < \infty$, $W^{j,p}(G)$ the space defined in Section 1 consisting of all u in $L^p(G)$ whose distribution derivatives of order $\leqq j$ also lie in $L^p(G)$. We wish to represent the conjugate space $(W^{j,p}(G))^*$ of $W^{j,p}(G)$ as a space whose elements are distributions with the functionals being given by the distribution pairing. That this is possible follows from the following Lemma:

LEMMA 9 (Calderon). *Let G be a firmly regular precompact open subset of E^n, $j \geqq 0$, $1 < p < \infty$. Then:*

(a) *The restriction mapping K of $W^{j,p}(E^n)$ into $W^{j,p}(G)$, with $Ru = u|_G$, is onto and there exists a bounded linear mapping J of $W^{j,p}(G)$ into $W^{j,p}(E^n)$ such that K is the identity map of $W^{j,p}(G)$.*

(b) *$R(G)$, the restrictions to G of the functions of $C^\infty{}_c(E^n)$, is dense in $W^{j,p}(G)$.*

(c) *Let $W^{-j,q}(E^n)$ be the conjugate space of $W^{j,p}(E^n)$ realized as a* *space of distributions in* E^n, $(q = p(p-1)^{-1})$. *Then the mapping* K^* *maps* $(W^{j,p}(G)^*)$ *injectively on a linear subset of distributions having their support in* $\bar{G}$.

Remark. If $W^{j,p}{}_0(G)$ is the closed subspace of $W^{j,p}(G)$ which is the closure in the latter space of the subset $C^\infty{}_0(G)$, we define $W^{-j,q}(G)$ $(q = p(p-1)^{-1})$ as the conjugate space of $W^{j,p}{}_0(G)$ realized in the obvious way as a family of distributions on G. In the case in which G coincides with E^n, $W^{j,p}{}_0(E^n)$ coincides with $W^{j,p}(E^n)$ and we obtain the space $W^{-j,q}(E^n)$ considered in Lemma 9.

Proof of Lemma 9. We begin by constructing a finite covering of G by open sets $\{N_k\}$ such that all but one are neighborhoods of boundary points satisfying the cone condition in the definition of the firm regularity of G and the last is an open set with compact closure in G. If $\{\eta_k\}$ is an infinitely differentiable partition of unity corresponding to the covering $\{N_k\}$, then in order to extend an element u of $W^{j,p}(G)$ to an element of $W^{j,p}(E^n)$, (or more precisely, in order to define Ku), it suffices to extend each of the functions $u_k = \eta_k u$, (or again more precisely, to set $Ku = \sum_k Ku_k$). We remark that part (a) of Lemma 9 implies (b) and (c), so that by the preceding construction, we can restrict ourselves to the proof of (a) for functions u which either have compact support in a fixed compact subset of the interior of G or in a neighborhood of the boundary satisfying the Sobolev cone condition for a cone with fixed axis. Such functions having compact support in G can be extended in a norm-preserving way by setting them zero outside of G, it suffices to consider only the case where the support of u lies in a suitable neighborhood of a boundary point.

Let u be a function in $W^{j,p}(G)$ vanishing outside a compact subset of a neighborhood N such that for every point x of $N \cap G$, the cone C_x given by

$$C_x = \{y : y = x + t\xi, 0 \leqq t < h, |\xi| = 1, |\xi - \xi_0| < h\}$$

lies in G for fixed h and ξ_0 with $h > 0$, $|\xi_0| = 1$. We choose a non-negative C^∞ function $h(\xi)$ on the unit sphere having compact support in the set $\{\xi : |\xi - \xi_0| < h\}$ and another function $k(t) \in C^\infty(R')$ with $k(t) = 1$ for $t \leqq d_0$, $k(t) = 0$ for $t \geqq h - d_0$, $d_0 > 0$.

Suppose first that u lies in C^j on some open set containing C_x in interior. Fixing ξ on the unit sphere S^{n-1} with $|\xi - \xi_0| < h$, we integrate by parts to obtain the following equation:

$$\begin{aligned}\int_0^h k(t)t^{j-1}\frac{d^j}{dt^j}u(x+t\xi)dt&=(-1)^{j-1}\int_0^h\frac{d^{j-1}}{dt^{j-1}}\{k(t)t^{j-1}\}\cdot\\ \cdot\{\frac{d}{dt}u(x+t\xi)\}dt&=(-1)^{j-1}\{-(j-1)!\,u(x)-\\ &\int_0^h\frac{d^j}{dt^j}[k(t)t^{j-1}]u(x+t\xi)dt\}.\end{aligned}\tag{6.1}$$

If we equate the first and third terms of equation (6.1) and solve for $u(x)$, expanding the derivatives with respect to t and setting

$$s(t)=-[(j-1)!]^{-1}\frac{d^j}{dt^j}[k(t)t^{j-1}]$$

we obtain

$$\begin{aligned}u(x)=\int_0^h k(t)t^{j-1}(\sum_{|\alpha|=j}C_\alpha\xi^\alpha(D^\alpha u)(x+\xi))dt\\ +\int_0^h s(t)u(x+t\xi)dt.\end{aligned}\tag{6.2}$$

The kernel $s(t)$ in equation (6.2) lies in C^∞ and vanishes outside of $d_0\leqq t\leqq h-d_0$. Multiplying eq. (6.2) by $h(\xi)dw(\xi)$, (where $dw(\xi)$ is the $(u-1)$-dimensional area element on S^{n-1}) and integrating over S^{n-1} (as we can do since h has its support in $\{\xi:|\xi-\xi_0|<h\}$), we obtain

$$\begin{aligned}C_0u(x)=\int\int k(t)t^{j-1}h(\xi)\{\sum_{|\alpha|=j}C_\alpha\xi^\alpha(D^\alpha u)(x+t\xi)dtdw(\xi)\\ +\int\int s(t)h(\xi)u(x+t\xi)dtdw(\xi),\ (C_0=\int_{S^{n-1}}h(\xi)d\xi\neq 0).\end{aligned}\tag{6.3}$$

If we introduce the n-variable y as variable of integration in eq. (6.3) with $dy=t^{n-1}dtdw(\xi)$, $y=t\xi$, and set

$$\begin{aligned}S(y)&=\frac{1}{C_0}s(|y|)h(y/|y|)|y|^{1-n}\\ K_\alpha(y)&=\frac{1}{C_0}k(|y|)|y|^{j-n}h(y/|y|)C_\alpha\xi^\alpha_y,\ (|\alpha|=j,\ \xi_y=y/|y|),\end{aligned}\tag{6.4}$$

we find that for $x\in N\cap G$

$$u(x)=\int S(y)u(x+y)dx+\sum_{|\alpha|=j}\int K_\alpha(y)(u)(x+y)dy$$

Since the functions S and the K_α have their support in the cone $\{y|y=t\xi,\ 0\leqq t<h,\ |\xi|=1,\ |\xi-\xi_0|<h\}$, the convolutions in equation (6.5) are well defined for $x\in N\cap G$.

We now extend the representation of u by eq. (6.5), which is essentially due to Sobolev [12] and which we therefore shall call the Sobolev representation, to all functions u in $W^{j,p}(G)$ having compact supports in $N \cap \bar{G}$. If we wish to establish eq. (6.5) almost everywhere in x in $N \cap G$, it suffices to show that it holds a.e. in x for x lying in a compact subset K of $N \cap G$. For $x \in K$, however, and y in the support of S or K_α, $u(x+y)$ and $(D^\alpha u)(x+y)$ have their support in a fixed compact subset K_1 of G containing K. We may approximate u on a neighborhood G' of K_1 in G in $W^{j,p}(G')$ by a sequence $\{u_t\}$ of functions in $C^\infty(G')$. For each of the u_t, the Sobolev representation (6.5) is valid for all x in K. Both sides of (6.5) are continuous mappings of $W^{j,p}(G')$ into $L^p(K)$, however, and by continuity it follows that eq. (6.5) must hold a.e. in K for an arbitrary u in $W^{j,p}(G)$. Hence eq. (6.5) holds a.e. in x on $N \cap G$ for all u in $W^{j,p}(G)$ in the compact support in $N \cap \bar{G}$.

Having the Sobolev representation (6.5) for u, it has been remarked by Calderon that the extension of u to an element of $W^{j,p}(E^n)$ follows very simply. We may assume without loss of generality that the support of u lies in a fixed compact subset K_2 of $N \cap \bar{G}$. Choose $\rho \in C^\infty{}_c(E^n)$ having compact support in N and equal to 1 on K_2, and set

$$(6.6)\qquad \begin{aligned} v(x) &= \begin{cases} u(x), & x \in G \\ 0 & , x \in G \end{cases} \\ v_\alpha(x) &= \begin{cases} (D^\alpha u)(x), & x \in G \\ 0 & , x \in G, \quad |\alpha| = j. \end{cases} \end{aligned}$$

Then v, v_α lie in $L^p(E^n)$ and we verify immediately from eq. (6.5) that for almost all x in $N \cap G$

$$(6.7)\qquad \begin{aligned} u(x) - u_1(x) &= \rho(x) \int_{E^n} S(y) v(x+y) \\ &\quad + \sum_{|\alpha|=j} \rho(x) \int K_\alpha(y) v_\alpha(x+y)\, dy. \end{aligned}$$

However, by the Calderon-Zygmund singular integral theorem, eq. (6.7) defines a function u_1 lying in $W^{j,p}(E^n)$ because $\rho \in C^\infty{}_c(E^n)$ and each K_α is C^∞ on the unit sphere S^{n-1} and has a singularity of the form $k_\alpha(\xi_y)\,|\,y\,|^{j-n}$. Moreover, u_1 as given by eq. (6.7) coincides with u a.e. in G. We let u_1 be the required extension of u and Lemma 9 is proved.

In the approximation arguments of this Section, we shall take our spaces F as $W^{j,p}(G)$, and for firmly regular G, we shall have $F^* \subset W^{-j,q}{}_c(E^n)$ with $q = p(p-1)^{-1}$.

To give a concise statement of our aprroximation theorems for non-elliptic differential operators, we shall employ the following definition:

DEFINITION. *Let A be a partial differential operator of order r with C^∞ coefficients on an open set G of E^n, G_1 and G_2 open subsets of G with $\bar{G}_1$ compact in G_2, $\bar{G}_2$ compact in G. Then the pair (G_1, G_2) is said to be an approximation pair with respect to A on G if the following two conditions hold:*

(1) *If $u \in \mathcal{E}'(G)$ has support in $\bar{G}_2$ and $A'u$ has its support in $\bar{G}_1$, then $u \in C^\infty(G - \bar{G}_1)$.*

(2) *If $u \in C^\infty(G - \bar{G}_1)$ u has its support in $\bar{G}_1$ and $A'u$ has its support in $\bar{G}_1$, then u must have its support in $\bar{G}_1$.*

An obvious remark is that we can combine (1) and (2) of the definition simply in the one condition:

(1)′ *If $u \in \mathcal{E}'(G)$ has its support in $\bar{G}_2$ with $A'u$ having its support in $\bar{G}_1$, then u must have its support in $\bar{G}_1$.*

We have separated this last condition into two parts in order to emphasize the fact that in the applications, the two parts are likely to be verified separately. In the terminology of Hörmander [8], condition (1) is the condition on the continuation of singularities while condition (2) is of the type of the uniqueness of the Cauchy problem in the small with respect to some class of non-characteristic surfaces.

LEMMA 10. *For $p = 2$, a pair of open subsets (G_1, G_2) is an approximation pair for the differential operator A in either of the following cases:*

(i) *A is hyperbolic and there exists a finite collection of continuous families of space-like hypersurfaces $\{S_t, 0 \leqq t \leqq 1\}$, which together sweep out $G - \bar{G}_1$.*

(ii) *A' is an operator of principal normal type (in the sense of Hörmander [8]) and there exists a finite number of families of surfaces $\{S_t\}$ continuous in t which sweep out $G - \bar{G}_1$, where S_t is the level surface of a function $\psi_t(x)$ which is strongly pseudo-convex with respect to A'.*

The proof of Lemma 10 follows from the proof of conditions (1) and (2), as given by Leray [10] for hyperbolic equations and by Hörmander [8] for operators of principal normal type with strongly pseudo-convex level surfaces.

THEOREM 6. ***Let A be a partial differential operator of order r with C^∞*** *coefficients on the open subset G of E^n, G_1 and G_2 firmly regular open subsets* *of G with $\bar{G}_1 \subset G_2$, G_2 contained and compact in G, the pair (G_1, G_2) forming an approximation pair with respect to A on G. Suppose furthermore that if $u \in W^{-j,q}{}_c(G)$ and $A'u \in W^{-j,q}{}_0(G)$ for $j \geqq 0$, $1 < q < +\infty$, then u must lie in $W^{r-j-1,q}{}_0(G)$. Let $p = q(q-1)^{-1}$.*

Let S be the family of solutions u of the equation $Au = 0$ in $W^{j,p}(G_2)$, S_1 the family of solutions u of the equation $Au = 0$ in G_1 lying in $W^{j,p}(G_1)$. Then the restrictions of S to G_1 are dense in S_1 in the norm of $W^{j,p}(G_1)$.

Proof of Theorem 6. The result will follow from Theorem 1 if we can verify the conditions:

(i) The domain of the minimal operator of A' in $W^{-j,q}(G_2)$ lies in the subset $W^{r-j-1}(G_2)$ with a compact imbedding map in $W^{-j,q}(G_2)$.

(ii) If u is a distribution in $W^{-r,q}(G_2) \subset W^{-r,q}{}_0(G)$ for which $A'u = 0$ on $G - \bar{G}_1$, then u lies in $C^\infty(G - \bar{G}_1)$.

(iii) If u lies in $C^\infty(G - \bar{G}_1)$, u has its support in $\bar{G}_2$ and $A'u = 0$ on $G - \bar{G}_1$, then u must be zero on $G - \bar{G}_1$.

(iv) Every u in $W^{-j,q}(G_2)$ for which $A'u$ also lies in $W^{-r,q}(G_2)$ with u having its support in $\bar{G}_1$, must be the limit in $W^{-r,q}(G_1)$ of a sequence of testing functions with compact support in G_1.

The condition (i), however, follows from the hypothesis of Theorem 6 since the domain of the minimal operator of A' in $W^{-j,q}(G_2)$ has a natural imbedding in the domain of the minimal operator of A' in $W^{-j,q}{}_0(G)$. Conditions (ii) and (iii) follows from the assumption that (G_1, G_2) is an approximation pair for A on G. Condition (iv) follows from the hypothesis that the domain of the minimal operator of A' in $W^{-j,q}(G_2)$ lies in $W^{r-1-j,q}(G_2)$ together with an obvious variant of Lemma 8. Thus the proof of Theorem 6 is complete.

COROLLARY TO THEOREM 6. *The conclusions of Theorem* 6 *follow for $p = 2$ under either of the hypotheses of Lemma* 10 *on the operator A.*

Proof of the Corollary. We need only remark that the traditional hypothesis employed in Theorem 6, namely that if u and $A'u$ both lie in $W^{-j,2}{}_c(G)$ then u must lie in $W^{r-1-j,2}{}_0(G)$, follows from results of Leray [10] and Hörmander [8] in the two cases. In case of operators of principal normal type, we must assume in addition that there exists a single function $\psi(x) \in C^2(G)$ which is pseudo-convex with respect to A at each point of a precompact neighborhood of $\bar{G}_2$ in G.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] H. Beckert, "Eine bemerkenswerte Eigenschaft der Lösungen des Dirichletschen Problems bei linearen elliptischen Differentialgleichungen," *Mathematische Annalen*, vol. 139 (1959), pp. 255-264.

[2] F. E. Browder, "On the spectral theory of elliptic differential operators I," *Mathematische Annalen*, vol. 142 (1961), pp. 22-130.

[3] ———, "Functional analysis and partial differential equations, II," *Mathematische Annalen*, vol. 145 (1962), pp. 81-226.

[4] ———, "Approximation in uniform norm by solutions of elliptic differential equations," *Bulletin of the American Mathematical Society*, vol. 67 (1961), pp. 400-404.

[5] ———, "Functional analysis and partial differential equations, III," *to appear.*

[6] A. P. Calderon, "Uniqueness in the Cauchy problem for partial differential equations," *American Journal of Mathematics*, vol. 80 (1958), pp. 16-36.

[7] K. O. Friedrichs, "On the differentiability of solutions of linear elliptic differential equations," *Communications on Pure and Applied Mathematics*, vol. 6 (1953), pp. 299-326.

[8] L. Hörmander, *Differential Equations with Non-Singular Characteristics*, Lecture Notes, American Mathematical Society Summer Conference Functional Analysis at Stanford, 1961.

[9] P. Lax, "A stability theory of abstract differential equations and its applications to the study of local behaviors of solutions of elliptic equations," *Communications on Pure and Applied Mathematics*, vol. 8 (1956), pp. 747-766.

[10] J. Leray, *Hyperbolic Equations*, Institute for Advanced Study, 1953.

[11] B. Malgrange, "Existence et approximation des solutions des équations aux derivées partielles et des équations de convolutions," *Ann. Inst. Fourier*, vol. 6 (1956), pp. 271-355.

[12] S. N. Mergelyan, "Uniform approximation to functions of a complex variable," *Usphi Mat. Nauk* (*N. S.*), vol. 7 (1952), pp. 31-122. (American Mathematical Society Translations, Series I, No. 101).

[13] S. L. Sobolev, *Some Applications of Functional Analysis to Mathematical Physics*, Leningrad, 1950.

SOME RESULTS ON ANALYTIC ITERATION AND CONJUGACY.*

By BENJAMIN MUCKENHOUPT.

1. Introduction. Given an analytic function $f(z)$ with $f(0)=0$, let $f_n(z)$ be defined for integral n by $f(f_{n-1}(z))=f_n(z)$ and $f_1(z)=f(z)$. One problem of iteration is to solve functional equations of the type $g_m(z)=f(z)$ where $f(z)$ is given or to assign meaning to $f_t(z)$ for t not an integer. Recent papers by Erdos and Jabotinsky [2] and Baker [1] have investigated this problem for $f'(0)=1$.[1] If $f(z)=z+a_nz^n+\cdots$, where $a_n\neq 0$, they considered the formal series $f_t(z)=z+ta_nz^n+\cdots$ where the rest of the terms are uniquely determined by the formal equation $f_t(f(z))=f(f_t(z))$. Baker showed that if $f(z)=e^z-1$, then $f_t(z)$ has positive radius of convergence only if t is an integer. Erdos and Jabotinsky showed that given a function $f(z)$ with $f'(0)=1$, $f_t(z)$ will either have a positive radius of convergence for all complex t or only for real t of one dimensional measure 0 and complex t of two dimensional measure 0. Furthermore, which is the case is determined by whether a certain associated series has a positive radius of convergence.

Given analytic functions $f(z)$ and $g(z)$ with $f(0)=g(0)=0$, one problem of conjugacy is that of deciding whether an analytic function $\phi(z)$ exists such that $\phi(0)=0$, $\phi'(0)\neq 0$, and $\phi^{-1}(f[\phi(z)])=g(z)$ in some neighborhood of the origin. This is closely related to the iteration problem and in several cases a solution of the iteration problem solves the conjugacy problem. When $|f'(0)|=|g'(0)|=1$, however, this is not true.

In this paper use will be made of the results in [1] and [2] to study the iteration and conjugacy problems for certain classes of functions. Section 3 considers the question of conjugacy of formal series with constant term zero and linear coefficient one while Section 4 solves the conjugacy problem for analytic functions of this type that have all their iterates. Section 5 considers the formal aspects of the fact that some of these functions have iterates in addition to those of the type considered before while Section 6 uses the

---

* Received November 6, 1961.

[1] For a treatment of the other cases see: for $|f'(0)|\neq 0$ or 1, [6] p. 58, for $f'(0)=0$, [4] p. 154, for $|f'(0)|=1$ but $f'(0)$ not a root of unity, [9] p. 608, for $|f'(0)|=1$ and $f'(0)$ a root of unity, Section 7 of this paper. [11] contains an approach to the n dimensional analogue.

result of Section 4 to prove that these iterates exist if the others exist. Finally, Section 7 treats the problems of conjugacy and iteration for functions with first derivative a root of unity.

2. Basic definitions and lemmas. Throughout this paper all functions will be assumed to be analytic and to vanish at the origin. Formal series will all have complex coefficients and no constant terms. To simplify notation $f(g(z))$ will be written $fg(z)$. Since the first n terms of $fg(z)$ depend only on the first n terms of $f(z)$ and $g(z)$, $fg(z)$ has meaning whether the series $f(z)$ and $g(z)$ have positive radius of convergence or not. Given $f(z)$, $f^{-1}(z)$ is the series such that $ff^{-1}(z) = z$, and if $f(z) = \sum_{j=1}^{\infty} a_j z^j$, then $f'(z) = \sum_{j=1}^{\infty} j a_j z^{j-1}$. The series $f'(z)$ may have a non zero constant term; this will be the only exception to the statement at the beginning of this paragraph.

Two series $f(z)$ and $g(z)$ are permutable if $fg(z) = gf(z)$.

Given a function $f(z) = z + a_n z^n + \cdots$ with $a_n \neq 0$, $f_t(z)$ is the unique series satisfying $ff_t(z) = f_t f(z)$ and having the form $z + t a_n z^n + \cdots$. The uniqueness follows from comparing the $k+n-1$-st terms of $ff_t(z)$ and $f_t f(z)$; they determined successively the k-th terms of $f_t(z)$. Furthermore, as stated in [2] the coefficient of z^j in $f_t(z)$ is a polynomial in t of degree at most $j-1$.

Given a series $f(z)$ with linear coefficient one, the associated series $L(z)$ is the formal solution of the equation $Lf(z) = f'(z)L(z)$ with its first non zero term equal to the second non zero term of $f(z)$.

Erdos and Jabotinsky showed in [2] that the first k coefficients of either a series or its associated series determine the first k coefficients of the other. In addition, the associated series has a positive radius of convergence if and only if all the series $f_t(z)$ have positive radius of convergence.

A function $f(z)$ with $f'(0) = 1$ is in the class C if its associated series has a positive radius of convergence.

LEMMA 1. *If* $f(z) = z + z^n + bz^{2n-1} + \cdots$, *its associated series.* $L(z) = z^n + (b - \frac{1}{2}n)z^{2n-1} + \cdots$.

This follows by comparing terms in the definition of $L(z)$.

LEMMA 2. *If* $f(z)$ *has the associated series* $L(z)$, *then* $\phi^{-1}f\phi(z)$ *has the associated series* $\dfrac{L\phi(z)}{\phi'(z)}$.

This can be verified by observing that the given series satisfies the necessary equation and has the proper first term.

3. Formal conjugacy considerations.

LEMMA 3. *Given a formal power series of the form* $f(z)=z+a_nz^n+\cdots, a_n\neq 0$, *there exists a polynomial* $\phi(z)$ *and a number* b *such that* $\phi^{-1}f\phi(z)=z+z^n+bz^{2n-1}+\cdots$.

Let $\phi(z)=\sum_{k=1}^{n-1}c_kz^k$ where the c_k's are to be determined. No matter what choice is made for the c_k's, the first $n-1$ terms of $f\phi(z)$ and $\phi(z+z^n+bz^{2n-1}+\cdots)$ will be the same. The n-th coefficients of these series are respectively $a_nc_1{}^n$ and c_1. Therefore, proper choice of $c_1\neq 0$ will make these terms equal. For $1<k\leqq n-1$, the $n+k-1$ st coefficients of $f\phi(z)$ and $\phi(z+z^n+bz^{2n-1}+\cdots)$ are respectively $na_nc_1{}^{n-1}c_k+$ (terms with lower c_j's) and kc_k+ (terms with lower c_j's). Since $a_nc_1{}^{n-1}=1$, the c_k's may be chosen successively to make all these coefficients equal. With this choice the first $2n-2$ terms of $f\phi(z)$ and $\phi(z+z^n+bz^{2n-1}+\cdots)$ are equal so that the first $2n-2$ terms of $\phi^{-1}f\phi(z)$ are $z+z^n$.

THEOREM 1. *Two power series with linear coefficients one are formally conjugate if and only if their second non zero terms are of the same order* n *and their reduced forms given by Lemma 3 have the same* $2n-1$*-st coefficients.*[2]

Because of Lemma 3 the sufficiency of this condition for conjugacy can be proved by showing that if

$$f(z)=z+z^n+bz^{2n-1}+\sum_{k=2n}^{\infty}c_kz^k \text{ and } g(z)=z+z^n+bz^{2n-1}+\sum_{k=2n}^{\infty}d_kz^k,$$

then $f(z)$ and $g(z)$ are conjugate. Let $h(z)=z+\sum_{k=n+1}^{\infty}e_kz^k$ where the e_k's are to be determined. For $j<2n$ the j-th terms of $hf(z)$ and $gh(z)$ are identical regardless of the choice of the e_k's. For $j\geqq 2n$ the j-th terms of $hf(z)$ and $gh(z)$ are respectively $e_j+(j-n+1)e_{j-n+1}+$ (terms in lower e_k's) and $e_j+ne_{j-n+1}+$ (terms in lower e_k's). Thus, proper successive choice of e_{j-n+1} wil make $hf(z)=gh(z)$ and prove the desired conjugacy.

To prove the necessity of the condition it will be shown that if $f(z)=z+z^n+bz^{2n-1}+\cdots$, $g(z)=z+z^m+cz^{2m-1}+\cdots$ and these series are conjugate, then $m=n$ and $b=c$. Let $h(z)=\sum_{k=1}^{\infty}d_kz^k$, $d_1\neq 0$, be the series such that $hf(z)=gh(z)$. If $m<n$, comparing m-th coefficients of $hf(z)$ and $gh(z)$ gives $d_m=d_m+d_1{}^m$. A similar contradiction is obtained by

---

[2] A general solution of the formal conjugacy problem that would include Theorem 1 and the formal part of Theorem 5 was announced in [3]. The statement and proof have not yet appeared.

assuming $m > n$. Therefore $m = n$. Comparing n-th terms of $hf(z)$ and $gh(z)$ then gives $d_1 + d_n = d_n + d_1{}^n$. The k-th terms for $n < k < 2n-1$ of these two series are respectively $d_k + (k-n+1)d_{k-n+1} +$ (terms with at least one d_j, $1 < j < k-n+1$) and $d_k + nd_1{}^{n-1}d_{k-n+1} +$ (terms with at least one d_j, $1 < j < k-n+1$). These show successively that $d_j = 0$ for $2 < j < n-1$. Then comparing $2n-1$-st terms of the two series shows that $d_1 b + nd_n + d_{2n-1} = d_{2n-1} + nd_n + d_1 c$. From this $b = c$.

4. Conjugacy of functions with first derivative 1. Although Theorem 1 gives a solution to the formal conjugacy problem and consequently a necessary condition for conjugacy of functions, the condition is not sufficient. The function $f(z)$ which has $\frac{z^2}{2} - \frac{z^3}{12}$ as its associated function has the first three terms $z + \frac{z^2}{2} + \frac{z^3}{6}$ by Lemmas 1 and 2. By the result in [2] mentioned previously, $f_t(z)$ has positive radius of convergence for all t. If there were a $\phi(z)$ with positive radius of convergence such that $\phi^{-1}f\phi(z) = e^z - 1$, then $\phi^{-1}f_{\frac{1}{2}}\phi(z) = g(z)$ would satisfy $gg(z) = e^z - 1$ contrary to Baker's result that $e^z - 1$ has no fractional iterates.

The example above suggests an additional obvious necessary condition for conjugacy of two functions: their iterates of the same order must both have positive radius of convergence or both must have radius of convergence zero. In general, this would be difficult to verify, and it is not clear whether even this is sufficient. However, for the most interesting case the following theorem gives a complete solution.

THEOREM 2. *If $f(z)$ and $g(z)$ are in the class C and are formally conjugate, they are conjugate.*

By Lemma 3 $f(z)$ and $g(z)$ are both conjugate to functions of the form $z + z^n + bz^{2n-1} + \cdots$, and by Theorem 1 the resulting values of n and b are the same for both functions. Consequently, it will be sufficient to prove this theorem for functions of this form.

Let $f(z)$ and $g(z)$ be two functions of this form that are in the class C. Let their associated functions be $L(z)$ and $M(z)$ respectively. Using Lemma 2 the problem reduces to finding a function $\phi(z)$ such that $\frac{L\phi(z)}{\phi'(z)} = M(z)$. Let $\phi(z) = z + yz^n$. The equation then reduces to

$$\frac{dy}{dz} = \frac{L(z+yz^n) - M(z) - nyz^{n-1}M(z)}{z^n M(z)}.$$

Since by Lemma 1 both $L(z)$ and $M(z)$ have the form $z^n + (b - \frac{1}{2}n)z^{2n-1} + \cdots$, the expression on the right side of the differential equation is analytic in a neighborhood of $y = 0$ and $z = 0$. Consequently, by Cauchy's existence theorem for differential equations,[3] there is an analytic function $y(z)$ which satisfies this differential equation. This gives a function $\phi(z)$ which satisfies the original equation.

5. Additional formal fractional iterates. Given a function $f(z)$ with $f'(0) = 1$, there may be more solutions to $g_m(z) = f(z)$ than just $f_{1/m}(z)$. For example, if $f(z) = \dfrac{z}{\sqrt{1-2z^2}}$, then $f_{\frac{1}{2}}(z) = \dfrac{z}{\sqrt{1-z^2}}$ and $-f_{\frac{1}{2}}(z)$ both satisfy the equation $gg(z) = f(z)$. Since any solution, $g(z)$, of an equation $g_m(z) = f(z)$ will satisfy $fg(z) = gf(z)$, necessary conditions for the existence of such other functions can be obtained by considering formally the equation $fg(z) = gf(z)$.

THEOREM 3. *If $f(z)$ is a series of the form $z + a_n z^n + \cdots$ where $a_n \neq 0$ and a is a primitive $n-1$-st root of* 1, *there exists a formal series $r(z)$ such that the linear coefficient of $r(z)$ is a, $r_{n-1}(z) = z$, and $f_t r(z) = r f_t(z)$ for all t. Furthermore, if $f_t g(z) = g f_t(z)$ formally for some $t \neq 0$, then either $g(z) = r_j f_u(z)$ where j is an integer and u a complex number or* $g(z) \equiv 0$.

By Lemma 3 and Theorem 1 there is a series $\phi(z)$ such that $\phi^{-1}f\phi(z) = z + z^n + bz^{2n-1}$. Let $q(z) = az$. Then since $q(z)$ permutes with $\phi^{-1}f\phi(z)$, $q(z)$ and $\phi^{-1}f_k\phi(z)$ permute for any positive integer k. Since $q(z)$ can permute only with a series that has all terms not of order $1 + (n-1)j$, j an integer, equal to zero, and since the coefficients of the terms of $\phi^{-1}f_t\phi(z)$ are polynomials in t, it follows that the coefficients of $\phi^{-1}f_t\phi(z)$ not of order $1 + (n-1)j$ are identically 0. Then since $q(z)$ permutes with all these series, $r(z) = \phi q \phi^{-1}(z)$ permutes with $f_t(z)$ for all t.

Now let $f_t(z) = z + cz^n + \cdots$, and $g(z) = \sum_{k=1}^{\infty} b_k z^k$. $c \neq 0$ since it is t times the n-th term of $f(z)$ and both these numbers differ from zero by hypothesis. Since $f_t g(z) = g f_t(z)$, comparing n-th terms gives $b_n + cb_1^n = b_1 c + b_n$. From this, b_1 is either 0 or an $n-1$-st root of one. The $n+k-1$-st terms of $f_t g(t)$ and $g f_t(z)$ are for $k > 1$ respectively $b_{n+k-1} + ncb_1^{n-1}b_k +$ (terms with lower b_j's) and $b_{n+k-1} + kb_k c +$ (terms with lower b_j's). If $b_1 = 0$, these successively determine all the b_k's as zero. Otherwire, since $b_1^{n-1} = 1$, these determine the b_k's uniquely except for b_n.

---

[3] [5] p. 347.

If $b_1 \neq 0$, there is an integer m such that $r_m(z)$ has first coefficient b_1 since $r(z)$ has first coefficient a primitive $n-1$-st root of unity. Since by proper choice of u, $f_u(z)$ can be given an arbitrary n-th term, there exists u such that $r_m f_u(z)$ has the same first and n-th terms as $g(z)$. Since g's other coefficients are uniquely determined by the fact that it permutes with $f_t(z)$ and since $r_m f_u(z)$ also permutes with $f_t(z)$, $g(z) = r_m f_u(z)$.

COROLLARY 1. *If $f(z)$ has linear coefficient* 1 *and $g(z)$ permutes with $f_t(z)$ for some $t \neq 0$, then $g(z)$ permutes with $f_t(z)$ for all t.*

COROLLARY 2. *If $f(z) = z + a_2 z^2 + \cdots$ and $a_2 \neq 0$, the only series that permute with $f(z)$ are the series $f_t(z)$ and* 0.

COROLLARY 3. *If $f(z) = a + a_n z^n + \cdots$, $a_n \neq 0$, and $g_m(z) = f(z)$ where m is an integer, then $g(z) = f_{1/m} r_k(z)$ where $r(z)$ is a series of the type mentioned in Theorem* 3 *and k is an integer such that mk is divisible by $n-1$.*

6. Convergence of the other iterates.

THEOREM 4. *If $f(z)$ is in the class C, then the series $r(z)$ described in Theorem* 3 *have positive radius of convergence.*

If the second non zero term of $f(z)$ is the n-th, Lemma 3 asserts the existence of a polynomial $\phi(z)$ and a constant b such that $\phi^{-1} f \phi(z) = z + z^n + b z^{2n-1} + \cdots$.

Now consider the function $g(z) = z + \sum_{i=n}^{\infty} b_i z^i$ with associated function $L(z) = z^n + (b + \frac{1}{2}n) z^{2n-1}$. It will be shown that the non zero terms of $g(z)$ are of order $1 + j(n-1)$, j an integer. If this were not true, there would be a first non zero coefficient, b_k, with k not of the form $1 + j(n-1)$. Then comparing $n+k-1$-st coefficients of $Lg(z)$ and $L(z)g'(z)$ would give $nb_k = kb_k$. This contradiction proves the desired result. Consequently, if a is a primitive $n-1$-st root of one and $q(z) = az$, then $qg(z) = gq(z)$.

Now $g(z)$ is in the class C since its associated function converges, and by Lemma 1 $g(z)$ has the form $z + z^n + bz^{2n-1} + \cdots$. By Theorem 2 there exists a function $h(z)$ such that $h^{-1} fh(z) = g(z)$. Since $q(z)$ permutes with $g(z)$, $hqh^{-1}(z)$ permutes with $f(z)$ and is therefore the desired $r(z)$.

As might be expected $r(z)$ may exist even if $f(z)$ is not in the class C. For example, $f(z) = \sqrt{e^{z^2} - 1}$ where the root is taken to make the linear coefficient 1, has all its even terms equal to zero. Consequently $r(z) = -z$ permutes with $f(z)$. $f(z)$ can have no iterates of non integral order, for if $f_t(z)$ has positive radius of convergence, so does $[f_t(\sqrt{z})]^2$ and this is an iterate of $e^z - 1$. Because of Lemma 4 in the next section, the general problem of the existence of $r(z)$ reduces to the problem of whether a function of the form $z + a_n z^n + \cdots, a_n \neq 0$, is conjugate to a function with terms not of order $1 + j(n-1)$ equal to zero. This, like the general conjugacy problem, remains an open question.

7. Iterates and conjugacy of functions with first derivative a root of unity.

This is a topic that seems to have been neglected in the literature although the results are simple and to a certain extent known. Since some of the results follow easily from the preceding theorems, it seems appropriate to include them here.

If $f(z) = a_1 z + \cdots$ and $a_1{}^n = 1$, there are two distinct cases depending on whether $f_n(z) = z$ or does not equal z. The two cases will be investigated in order.

LEMMA 4. *If* $f(z) = \sum_{k=1}^{\infty} a_k z^k$, *and* $f_n(z) = z$, *then there exists a function* $\phi(z)$ *such that* $\phi^{-1} f \phi(z) = a_1 z$ *in some neighborhood of the origin.*[4]

Let A be a disc about 0 so small that $f_j(z)$ is schlicht on A and $f_j(A)$ is convex for all integers j.[5] Let B be the union of all the sets $f_j(A)$. Then B is simply connected and $f(z)$ maps B one to one onto itself. Let $\phi(z)$ be the function that maps the unit circle onto B and leaves the origin fixed. Then $\phi^{-1} f \phi(z)$ maps the unit circle onto itself and leaves the origin fixed and as a result is equal to a constant times z. The constant must be a_1 so that $\phi^{-1} f \phi(z) = a_1 z$.

COROLLARY 4. *If* $f_n(z) = z$, *then* $g(z)$ *is conjugate to* $f(z)$ *if and only if* $g_n(z) = z$ *and the linear coefficient of* $g(z)$ *equals the linear coefficient of* $f(z)$.

COROLLARY 5. *If* $f_n(z) = z$, *then the equation* $g_m(z) = f(z)$ *has an infinite number of solutions for every integer* $m > 1$.

If $f(z) = a_1 z + \cdots$ and $\phi^{-1} f \phi(z) = a_1 z$, let $h(z)$ be any function that permutes with $a_1 z$, let b be an m-th root of a_1, and let $q(z) = bz$. Then $g(z) = \phi h q h^{-1} \phi^{-1}(z)$ satisfies the given functional equation.

---

[4] This result is contained in [10] pp. 159-161.

[5] To show that these domains can be made convex see [8] p. 224, ex. 5.

THEOREM 5. *If* $f(z) = a_1 z + \cdots$, a_1 *is a primitive* j*-th root of* 1 *and* $f_j(z) \neq z$, *then* $f(z)$ *is conjugate to* $g(z) = b_1 z + \cdots$ *if and only if* $a_1 = b_1$ *and* $f_j(z)$ *is conjugate to* $g_j(z)$.

If $f(z)$ and $g(z)$ are conjugate the conditions clearly hold. Conversely, if they hold let $g_j(z) = p(z)$. By hypothesis there exists a function $\phi(x)$ such that $\phi^{-1} f_j \phi(z) = p(z)$. Then the function $h(z) = \phi^{-1} f \phi(z)$ permutes with $p(z)$ since $h_j(z) = p(z)$. By Theorem 3 $h(z) = p_{1/j} r_k(z)$ formally where $r(z)$ is a series whose existence is asserted by Theorem 3. Using Theorem 3 again $g(z) = p_{1/j} r_m(z)$ formally since $g(z)$ permutes with $p(z)$. The series $r_k(z)$ and $r_m(z)$ must be the same since they have the same linear coefficients. Therefore, $h(z) = \phi^{-1} f \phi(z) = g(z)$ formally, and since all these series have positive radius of convergence, $f(z)$ is conjugate to $g(z)$.

Now let (a, b) denote the greatest common divisor of a and b.

THEOREM 6. *If* $f(z) = a_1 z + \cdots$, a_1 *is a primitive* j*-th root of one and* $b_j(z) = z + b_n z^n + \cdots$ *where* $b_n \neq 0$, *then* j *divides* $n - 1$. *Furthermore, the equation* $g_m(z) = f(z)$ *has* $(m, n - 1)$ *formal solutions if* $(m, n - 1)$ *divides* $(n - 1)/j$ *and none if* $(m, n - 1)$ *does not divide* $(n - 1)/j$. *If* $f_j(z)$ *is in theclass* C, *all the formal solutions of* $g_m(z) = f(z)$ *have positive radius of convergence.*

Let $h(z) = f_j(z)$ and let $r(z)$ be a series whose existence is asserted by Theorem 3. Then there exists an integer k such that $f(z) = r_k h_{1/j}(z)$ since $f(z)$ permutes with $h(z)$. If the linear coefficient of $r(z)$ is b, then $b^k = a_1$ whence $a_1{}^{n-1} = 1$. Therefore, a_1's order divides $n - 1$ as asserted. Now since $r(z)$ can have its linear coefficient, b, an arbitrary primitive $n - 1$-st root of one, choose it so that $b^{(n-1)/j} = a_1$.

If $g_m(z) = f(z)$ formally, then $g(z)$ permutes formally with $h(z)$. Then $g(z) = r_i h_{1/jm}(z)$ for some integer i. Comparing linear terms of $g_m(z)$ and $f(z)$ gives $b^{mi} = b^{(n-1)/j}$. This has the same values of i for solutions as the congruence $mi \equiv (n - 1)/j \pmod{n - 1}$. The congruence, however, has the number of solutions asserted.[6]

Because of Theorem 4 $f_j(z)$ being in the class C implies that $r(z)$ has positive radius of convergence; this along with the fact that $h_t(z)$ has positive radius of convergence for all t gives the desired existence.

RUTGERS, THE STATE UNIVERSITY.

---

[6] See for instance [7] p. 77.

REFERENCES.

[1] I. N. Baker, "Zusammensetzungen Ganzen Funktionen," *Mathematische Zeitschriften,* vol. 69 (1958), pp. 121-163.

[2] Paul Erdos and Eri Jabotinsky, "On Analytic iteration," *Journal d'Analyse Mathematique,* vol. 8 (1961), pp. 361-376.

[3] N. J. Fine and Bertram Kostant, "The group of formal power series under iteration," *Bulletin of the American Mathematical Society,* vol. 61 (1955), pp. 36-37.

[4] Gaston Julia, "Mémoire sur la Convergence des Series Formé avec les Itérées Successives d'une Fraction Rationnelle," *Acta Mathematica,* vol. 56 (1931), pp. 159-195.

[5] E. Goursat, *Cours d'Analyse Mathématique,* vol. II, Paris, 1911.

[6] H. Kneser, "Reele Analytischer Lösungen der Gleichung $\phi[\phi(z)] = e^z$ und Verwandter Funktionalgleichungen," *Journal für die Reine und Angewandte Mathematik,* vol. 187 (1949), pp. 56-67.

[7] Trygve Nagell, *Introduction to Number Theory,* John Wiley & Sons, 1951.

[8] Zeev Nehari, *Conformal Mapping,* McGraw-Hill, 1952.

[9] C. L. Siegel, "Iteration of analytic functions," *Annals of Mathematics,* vol. 43 (1942), pp. 607-612.

[10] ———, *Vorlesungen über Himmelsmechanik,* Springer-Verlag, 1956.

[11] S. Sternberg, "Infinite Lie groups and the formal aspects of dynamical systems," *Journal of Mathematics and Mechanics,* vol. 10 (1961), pp. 451-474.

THE CONFORMAL TRANSFORMATION GROUP OF A COMPACT RIEMANNIAN MANIFOLD.*

By SAMUEL I. GOLDBERG[1] and SHOSHICHI KOBAYASHI.

1. Introduction. Let M be a Riemannian manifold, $C_0(M)$ the largest connected group of conformal transformations of M and $I_0(M)$ the largest connected group of isometries of M. The purpose of the present paper is to prove the following

THEOREM 1. *Let M be a compact Riemannian manifold. If $C_0(M) \neq I_0(M)$, then there is no harmonic form of degree p, $0 < p < \dim M$, whose length is a non-zero constant.*

Since a harmonic form on a compact Riemannian manifold is invariant by $I_0(M)$, a harmonic form on a compact homogeneous Riemannian manifold is of constant length. As an immediate consequence of Theorem 1, we have the following

THEOREM 2. *Let M be a compact homogeneous Riemannian manifold. If $C_0(M) \neq I_0(M)$, then M is a rational homology sphere.*

We list known results in the same direction: (The present paper is self-contained and does not depend on any of the following results.)

A. *If M is a complete Riemannian manifold with parallel Ricci tensor, then $C_0(M) \neq I_0(M)$ implies that M is isometric to a sphere.* (Nagano [4], Nagano-Yano [5]).

B. *If M is a compact (almost) Kaehler manifold of real dimension > 2, then $C_0(M) = I_0(M)$.* (Lichnerowicz [3], p. 148, Goldberg [1, 2] Tachibana [6]).

C. *If M is a compact Riemannian manifold with negative constant curvature scalar, then $C_0(M) = I_0(M)$.* (Lichnerowicz [3], p. 134).

Note that B follows from Theorem 1, because the fundamental 2-form of a compact (almost) Kaehler manifold is a harmonic form with constant length.

---

* Received November 30, 1961.

[1] The first named author was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49 (638)-967.

Let M be a Riemannian manifold with metric tensor $ds^2 = \sum g_{ij} dx^i dx^j$. A transformation ϕ of M is said to be *conformal* if $\phi^*(ds^2) = f \cdot ds^2$, where f is a function on M. If f is a constant function, then ϕ is called *homothetic.* If f is identically equal to 1, ϕ is an *isometry.*

Since we are interested in connected groups, we can replace the hypothesis of Theorem 1 by the following. *Let M be a compact Riemannian manifold which admits an infinitesimal non-isometric conformal transformation.* We may also assume that M is orientable; if M is not orientable, we have only to take an orientable two-fold covering space of M.

2. Notations and formulae. For p-forms

$$\alpha = \frac{1}{p!} \sum \alpha_{i_1 \cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p} \text{ and } \beta = \frac{1}{p!} \sum \beta_{j_1 \cdots j_p} dx^{j_1} \wedge \cdots \wedge dx^{j_p}$$

(where the coefficients are assumed to be skew symmetric), we set

$$\langle \alpha, \beta \rangle = \frac{1}{p!} \sum g^{i_1 j_1} \cdots g^{i_p j_p} \alpha_{i_1 \cdots i_p} \beta_{j_1 \cdots j_p} \tag{1}$$

and

$$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle dV, \tag{2}$$

where dV is the volume element of M. $\langle\ ,\ \rangle$ and $(\ ,\ )$ are called the local and the global scalar products, respectively. The co-differential δ is the dual of d so that

$$(d\alpha, \beta) = (\alpha, \delta\beta). \tag{3}$$

Let $X = \sum \xi^i \frac{\partial}{\partial x^i}$ be an infinitesimal transformation and $\xi = \sum \xi_i dx^i$ the 1-form associated with X, i. e., $\xi_i = \sum g_{ij} \xi^j$. Let $\theta(X)$ denote Lie derivation with respect to X. For a differential form α, we have

$$\theta(X)\alpha = d \circ i(X)\alpha + i(X) \circ d\alpha, \qquad d \circ \theta(X)\alpha = \theta(X) \circ d\alpha. \tag{4}$$

where $i(X)$ is the interior product with respect to X. For the metric tensor, we have

$$\theta(X) \cdot ds^2 = \sum (\nabla_j \xi_i + \nabla_i \xi_j) dx^i dx^j, \tag{5}$$

where ∇ denotes covariant differentiation. An infinitesimal transformation X is conformal, if and only if

$$\nabla_j \xi_i + \nabla_i \xi_j = \lambda g_{ij}, \tag{6}$$

where λ is a function on M. It is homothetic, if and only if λ is a constant function, and it is isometric, if and only if $\lambda \equiv 0$. For a 1-form ξ, we have

$$\delta\xi = -\sum \nabla_i \xi^i. \tag{7}$$

Multiplying (6) by g^{ij} and contracting, we obtain $-2\delta\xi = n\lambda$. Hence, if X is conformal

$$\nabla_j \xi_i + \nabla_i \xi_j = -\frac{2}{n} \delta\xi \, g_{ij} \quad (n = \dim M). \tag{8}$$

Applying $\theta(X)$ to (1) with $\alpha = \beta$, we have, if X is conformal

$$\theta(X)\langle\alpha, \alpha\rangle = 2\langle\theta(X)\alpha, \alpha\rangle + \frac{2p}{n}\delta\xi\langle\alpha, \alpha\rangle. \tag{9}$$

Define $\bar{\theta}(X)$ to be the dual of $\theta(X)$, i.e., $(\bar{\theta}(X)\alpha, \beta) = (\alpha, \theta(X)\beta)$. From the second formula of (4) we obtain

$$\delta \circ \bar{\theta}(X)\alpha = \bar{\theta}(X) \circ \delta\alpha. \tag{10}$$

If X is conformal, then

$$\theta(X)\alpha + \bar{\theta}(X)\alpha = (1 - \frac{2p}{n})\delta\xi \cdot \alpha. \tag{11}$$

The proof of (11) will be given at the end of the paper. (This formula has already been used by Goldberg in [1, 2].)

3. Proof of Theorem 1. Let α be a harmonic form of degree p. We shall first prove

$$(\bar{\theta}(X)\alpha, \theta(X)\alpha) = 0. \tag{12}$$

Since α is closed, $\theta(X)\alpha = d \circ i(X)\alpha$ by (4). Since α is co-closed,

$$(\bar{\theta}(X)\alpha, \theta(X)\alpha) = (\delta \circ \bar{\theta}(X)\alpha, i(X)\alpha) = (\bar{\theta}(X) \circ \delta\alpha, i(X)\alpha) = 0.$$

From (11) and (12), we obtain (assuming that X is conformal)

$$\begin{aligned} (\theta(X)\alpha, \theta(X)\alpha) &= (\theta(X)\alpha + \bar{\theta}(X)\alpha, \theta(X)\alpha) \\ &= (1 - \frac{2p}{n})(\delta\xi \cdot \alpha, \theta(X)\alpha) \\ &= (1 - \frac{2p}{n}) \int_M \delta\xi \cdot \langle\alpha, \theta(X)\alpha\rangle dV. \end{aligned} \tag{13}$$

From now on we shall assume that α is not only harmonic but is of constant length, i. e., $\langle\alpha,\alpha\rangle$ is constant. Hence, $\theta(X)\langle\alpha,\alpha\rangle=0$. From (9) we obtain

$$\langle\theta(X)\alpha,\alpha\rangle=-\frac{p}{n}\delta\xi\langle\alpha,\alpha\rangle. \tag{14}$$

Substituting (14) into (13), we obtain

$$(\theta(X)\alpha,\theta(X)\alpha)=-(1-\frac{2p}{n})\frac{p}{n}\int_M(\delta\xi)^2\langle\alpha,\alpha\rangle dV. \tag{15}$$

Assume $0<2p\leqq n$. Then the right hand side of (15) is non-positive and the left hand side is non-negative. Hence, $\theta(X)\alpha=0$. By (14), $\delta\xi=0$ or $\alpha=0$. By (8), $\delta\xi=0$ implies that X is isometric. We proved therefore that if M admits an infinitesimal non-isometric conformal transformation X, then there is no harmonic form α of constant length and of degree p, $0<p\leqq n/2$. If α is a harmonic form of constant length and of degree $p>n/2$, then its adjoint $*\alpha$ is a harmonic form of constant length and of degree $n-p<n/2$. This completes the proof of Theorem 1.

4. Proof of (11). Applying $\theta(X)$ to (1), we obtain

$$\theta(X)\langle\alpha,\beta\rangle=\langle\theta(X)\alpha,\beta\rangle+\langle\alpha,\theta(X)\beta\rangle+\frac{2p}{n}\delta\xi\langle\alpha,\beta\rangle. \tag{16}$$

We also have

$$\theta(X)dV=-\delta\xi\, dV. \tag{17}$$

From (16) and (17), we obtain

$$\begin{aligned}\theta(X)(\langle\alpha,\beta\rangle dV)=\langle\theta(X)\alpha,\beta\rangle dV+\langle\alpha,\theta(X)\beta\rangle dV\\ +(\frac{2p}{n}-1)\delta\xi\langle\alpha,\beta\rangle dV.\end{aligned} \tag{18}$$

The integral of the left hand side of (18) over M vanishes because, for any n-form ω,

$$\int\theta(X)\omega=\int d\cdot i(X)\omega=0$$

by Stokes' formula. By integrating (18), we obtain

$$0=(\theta(X)\alpha,\beta)+(\alpha,\theta(X)\beta)-((1-\frac{2p}{n})\delta\xi\cdot\alpha,\beta).$$

Hence,

$$(\theta(X)\alpha+\bar{\theta}(X)\alpha,\beta)=((1-\frac{2p}{n})\delta\xi\cdot\alpha,\beta).$$

Since the above formula holds for all α and β, we obtain (11).

Remark. A compact symmetric space which is a rational homology sphere is isometric to a sphere, except for $SU(3)/SO(3)$ (Kostant: On invariant skew tensors, Proc. Nat. Acad. Sci., vol. 42 (1956), pp. 148-151).

UNIVERSITY OF ILLINOIS,
UNIVERSITY OF BRITISH COLUMBIA.

REFERENCES.

[1] S. I. Goldberg, "Groups of automorphisms of almost Kaehler manifolds," *Bulletin of the American Mathematical Society*, vol. 66 (1960), pp. 180-183.

[2] ———, "Groups of transformations of Kaehler and almost Kaehler manifolds," *Commentarii Mathematici Helvetici*, vol. 35 (1961), pp. 35-46.

[3] A. Lichnerowicz, *Géométrie des groupes de transformations*, Dunod, Paris (1958).

[4] T. Nagano, "The conformal transformation on a space with parallel Ricci tensor," *Journal of the Mathematical Society of Japan*, vol. 11 (1959), pp. 10-14.

[5] T. Nagano and K. Yano, "Einstein spaces admitting a one-parameter group of conformal transformations," *Annals of Mathematics*, vol. 69 (1959), pp. 451-461.

[6] S. Tachibana, "On almost analytic vectors in almost Kaehlerian manifolds," *Tohoku Mathematical Journal*, vol. 11 (1959), pp. 247-265.

ON SIEGEL MODULAR FORMS OF GENUS TWO.*[1]

By JUN-ICHI IGUSA.

I. Introduction. Since the publication of Siegel's *Einführung in die Theorie der Modulfunktionen n-ten Grades,* the theory has been enriched by several mathematicians. However, even in the case of degree two, the whole subject is rather misty compared with the clarity we have in the elliptic case. We are not talking here about the superstructure but the very foundation of the theory. As Siegel's treatment already shows, one of the fundamental objects of study is the graded ring of (finite sums of) modular forms. However, we know very little about this ring beyond the facts that it is finitely generated and that an operator Φ introduced by Siegel is almost an epimorphism of the graded rings of degree n to degree $n-1$ [3, 6]. Now, among modular forms, Eisenstein series (in the original sense of Siegel) are singled out by their importance in the analytic theory of quadratic forms [9]. We shall show that, *in the degree two case, every modular form is a polynomial of Eisenstein series of weight four, six, ten and twelve.* These four Eisenstein series are, of course, algebraically independent. Thus, we have a *complete structure theorem* of the ring and it gives answers to some well-known problems in this field, e. g. the dimension of the complex vector space of modular forms of a given weight and the structure of "Satake's compactification" of Siegel's fundamental domain (in the degree two case). We shall also determine the birational correspondence between the projective varieties associated with the graded ring of "even" projective invariants of binary sextics and with the graded ring of modular forms. In other words, we shall obtain explicit rational expressions for the three fundamental absolute invariants in terms of the four Eisenstein series.

We shall give an outline of our method. The results in AVM imply that the projective variety associated with the graded ring of even projective invariants of binary sextics is a compactification of the variety of moduli of curves of genus two. However, this projective variety does not contain the Siegel fundamental domain. In fact, those points of the Siegel fundamental domain representing products of elliptic curves are all mapped to one simple

---

* Received November 4, 1961.

[1] This work was partially supported by the Alfred P. Sloan Foundation.

point of the projective variety. The first thing we do is, therefore, to "blow up" this point to a two dimensional affine space to construct the Siegel fundamental domain explicitly. We note that the blowing up is not a monoidal transformation. At any rate, once we know the structure of the Siegel fundamental domain as an (incomplete) algebraic variety, we can very easily characterize "multi-canonical differentials" on the variety which correspond to modular forms whose weights are multiples of six. In this way, we are able to determine the structure of the graded ring of (finite sums of) such modular forms. Then, by taking its normalization in the field of fractions of modular forms, we get the graded ring of all modular forms and can prove that it is generated by four modular forms of weight four, six, ten and twelve. The relation between these modular forms and Eisenstein series will finally be obtained by comparing their Fourier expansions. In this paper, we shall use classical formulae on elliptic theta-functions due to Jacobi and well-known results on elliptic modular forms, i.e. those we can find in Hurwitz's papers, without specific references. We shall use, also, the following standard notations:

$\mathfrak{S}_n$ = Siegel upper-half plane of degree n, i.e. the variety of complex symmetric matrices of degree n with positive-definite imaginary parts

Γ_n = homogeneous modular group of degree n ($= Sp(n, \boldsymbol{Z})$) operating in $\mathfrak{S}_n$ as

$$\tau \to (a\tau + b)(c\tau + d)^{-1}$$

F_n = Siegel fundamental domain of degree n ($= \Gamma_n \backslash \mathfrak{S}_n$).

II. Blowing up of proj $\boldsymbol{C}[A, B, C, D]$.

1. Let A, B, C, D be the projective invariants of binary sextics of degree two, four, six, ten defined in AVM. If $\xi_1, \xi_2, \cdots, \xi_6$ are roots of a sextic

$$u_0 X^6 + u_1 X^5 + \cdots + u_6$$

and if we denote $\xi_j - \xi_k$ by (jk), the values of A, B, C, D at this sextic have the following irrational expressions

$$A(u) = u_0^2 \sum_{\text{fifteen}} (12)^2(34)^2(56)^2$$

$$B(u) = u_0^4 \sum_{\text{ten}} (12)^2(23)^2(31)^2(45)^2(56)^2(64)^2$$

$$C(u) = u_0^6 \sum_{\text{sixty}} (12)^2(23)^2(31)^2(45)^2(56)^2(64)^2(14)^2(25)^2(36)^2$$

$$D(u) = u_0^{10} \prod_{j<k} (jk)^2.$$

In general, if S is a finitely generated graded integral ring over the field $\boldsymbol{C}$ of complex numbers, we shall denote by $\operatorname{proj} S$ the (complex) projective variety associated with S.[2] We are interested in the following projective variety:

$$X = \operatorname{proj} \boldsymbol{C}[A, B, C, D].$$

We have shown in AVM that points of X which are not on $D = 0$ form the variety of moduli of curves of genus two, hence also of their jacobian varieties (with canonical polarizations) over $\boldsymbol{C}$. Now, the jacobian varieties can degenerate to products of ellipitic curves, and they form a two dimensional affine space. However, since the projective invariants B, C, D vanish simultaneously at sextics with triple roots, such abelian varieties are mapped to one point of X. We shall, therefore, try to blow up this point to get the two dimensional affine space of products of elliptic curves. We note that, since the points of X which are not on $A = 0$ form a three dimensional affine space, the point in question is, at any rate, simple on X. In fact, the following three absolute invariants

$$B/A^2 \qquad C/A^3 \qquad D/A^5$$

form a set of uniformizing parameters of X around this point. Therefore, we take a point

$$\tau = \begin{pmatrix} \tau_1 & \tau_{12} \\ \tau_{12} & \tau_2 \end{pmatrix}$$

of $\mathfrak{S}_2$ and try to expand the above three absolute invariants into power-series of $\epsilon = \tau_{12}$ (assuming that ϵ is small). The actual calculation was guided by the following observation.

Suppose that $\tau_{12} = 0$ in τ. Then, certainly τ corresponds to a product of elliptic curves whose Weierstrass invariants are $j(\tau_1)$ and $j(\tau_2)$. However, if we start from a product of elliptic curves, there exist infinitely many points in $\mathfrak{S}_2$, some not even satisfying $\tau_{12} = 0$, which correspond to the product and ϵ does not have any intrinsic meaning. Consider the subvariety of $\mathfrak{S}_2$ defined by the equation $\tau_{12} = 0$. Then, we can see by a simple matrix calcu-

---

[2] In Grothendieck's language, what we are defining is

$$\mathrm{Proj}(S)_C(C) = \mathrm{Hom}_{\mathrm{Spec}(C)}(\mathrm{Spec}(C), \mathrm{Proj}(S)).$$

Since we are assuming that S is finitely generated, this point-set has the unique structure of a projective variety, i. e. a closed subvariety of a projective space.

lation that elements of Γ_2 which keep this subvariety stable are those we can expect, i.e. of the following two types

$$\begin{pmatrix} a_1 & 0 & b_1 & 0 \\ 0 & a_2 & 0 & b_2 \\ c_1 & 0 & d_1 & 0 \\ 0 & c_2 & 0 & d_2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 0 & b_1 & 0 \\ 0 & a_2 & 0 & b_2 \\ c_1 & 0 & d_1 & 0 \\ 0 & c_2 & 0 & d_2 \end{pmatrix}.$$

By these elements, the parameter $\epsilon = \tau_{12}$ of points of a tubular neighborhood of the subvariety $\tau_{12} = 0$ undergoes the following transformation:

$$\epsilon \to \epsilon/(c_1\tau_1 + d_1)(c_2\tau_2 + d_2) \bmod \epsilon^2.$$

Therefore, if we take powers of ϵ and multiply elliptic modular forms in τ_1 and τ_2 of the corresponding weight, we can make it invariant modulo higher powers of ϵ. If we require, in addition, that the modular forms to be multiplied have no zeros, which is certainly a reasonable condition to get a good parameter, the smallest weight of the modular forms will be twelve. In this way, we can foresee that the parameter we need is $\delta(\tau_1)\delta(\tau_2)\epsilon^{12}$ up to some normalizing constant with

$$\delta(\omega) = e(\omega) \prod_{n=1}^{\infty} (1 - e(n\omega))^{24}$$

in which $e(\ )$ stands for $\exp(2\pi i\ )$ and ω is a point of $\mathfrak{S}_1$. This was a very discouraging conclusion because it forced us to calculate expansions of absolute invariants up to order twelve in ϵ. However, very fortunately, the blowing up is not monoidal and this complication on the part of the transformation brought about an unexpected simplification of the other.

2. As in the previous section, let τ be a point of $\mathfrak{S}_2$. Then, four column vectors of a two-by-four matrix $(\tau 1_2)$ generate a discrete subgroup of two dimensional complex vector space and the corresponding quotient group T is a complex torus carrying a positive divisor Θ satisfying $\deg(\Theta, \Theta) = 2$. Moreover, by a suitable translation in T, we can make Θ symmetric in the sense that it is stable under the transformation $z \to -z$. There exist sixteen such divisors and they are zeros of the following theta-functions:

$$\theta_{{}^t\mathfrak{g}{}^t\mathfrak{h}}(\tau, z) = \sum_{n} e[\tfrac{1}{2} \cdot {}^t(n + \mathfrak{g}/2)\tau(n + \mathfrak{g}/2) + {}^t(n + \mathfrak{g}/2)(z + \mathfrak{h}/2)]$$

in which $\mathfrak{g}$ and $\mathfrak{h}$ are column vectors with 0, 1 as their coefficients. On the other hand, if the point τ is not equivalent (with respect to Γ_2) to a point on $\tau_{12}=0$, we know that the divisor Θ is actually a non-singular curve of genus two and T is its jacobian variety [11]. In this case, the three lambdas which appear in what we called Rosenhain normal form of Θ in AVM can be expressed by theta-functions of zero argument

$$\theta_{{}^t\mathfrak{g}{}^t\mathfrak{h}}(\tau,0)=\theta_{{}^t\mathfrak{g}{}^t\mathfrak{h}}(\tau)=\theta_{{}^t\mathfrak{g}{}^t\mathfrak{h}}.$$

We have a choice of seven-hundred twenty expressions, and we shall use the following one

$$\lambda_1=(\theta_{1100}\theta_{1000}/\theta_{0100}\theta_{0000})^2 \qquad \lambda_2=(\theta_{1001}\theta_{1100}/\theta_{0001}\theta_{0100})^2$$

$$\lambda_3=(\theta_{1001}\theta_{1000}/\theta_{0001}\theta_{0000})^2$$

which is given by Rosenhain [7]. Conversely, if λ_1, λ_2, λ_3 are three complex numbers different from each other and from 0, 1, ∞, we can consider a period matrix belonging to the corresponding Rosenhain normal form. If we normalize the period matrix suitably in the form $(\tau 1_2)$, the three lambdas can be expressed as above with respect to this τ. We shall try to expand the lambdas into power-series of $\epsilon=\tau_{12}$. We observe that, in case neither (g_1,h_1) nor (g_2,h_2) is $(1,1)$, we have

$$\theta_{g_1g_2h_1h_2}\begin{pmatrix}\tau_1 & \epsilon\\ \epsilon & \tau_2\end{pmatrix}=\sum_{n=0}^{\infty}\frac{2^{2n}}{(2n)!}\,\frac{d^n}{d\tau_1{}^n}\theta_{g_1h_1}(\tau_1)\,\frac{d^n}{d\tau_2{}^n}\theta_{g_2h_2}(\tau_2)\cdot\epsilon^{2n}.$$

The verification is formal and straightforward. Therefore, we can express the coefficients of the expansions

$$\lambda_1=a_0+a_1\epsilon^2+a_2\epsilon^4+\cdots$$

$$\lambda_2=b_0+b_1\epsilon^2+b_2\epsilon^4+\cdots$$

$$\lambda_3=c_0+c_1\epsilon^2+c_2\epsilon^4+\cdots$$

in terms of elliptic theta-functions of zero argument and their derivatives. For instance, we have

$$a_0=b_0=c_0=\lambda(\tau_1)$$

in which

$$\lambda(\omega)=(\theta_{10}(\omega)/\theta_{00}(\omega))^4.$$

Now, we evaluate the projective invariants A, B, C, D at a sextic $X(X-1)(X-\lambda_1)(X-\lambda_2)(X-\lambda_3)$ and replace the three lambdas by their expansions in ϵ. Then, writing $\lambda=\lambda(\tau_1)$ for the sake of simplicity, we get the following expansions

$$
\begin{aligned}
A &= 6\lambda^2(1-\lambda)^2+4\lambda(1-\lambda)(1-2\lambda)\sum a_1\epsilon^2+\cdots\\
B &= 2\lambda^2(1-\lambda)^2(1-\lambda+\lambda^2)\sum(b_1-c_1)^2\epsilon^4\\
&\quad+4\lambda(1-\lambda)(1-2\lambda)\sum a_1(b_1-c_1)^2\epsilon^6\\
&\quad-2\lambda^2(1-\lambda)^2(1-2\lambda)\sum a_1\sum(b_1-c_1)^2\epsilon^6\\
&\quad+4\lambda^2(1-\lambda)^2(1-\lambda+\lambda^2)\sum(b_1-c_1)(b_2-c_2)\epsilon^6+\cdots
\end{aligned}
$$

$$
\begin{aligned}
C &= 4\lambda^4(1-\lambda)^4(1-\lambda+\lambda^2)\sum(b_1-c_1)^2\epsilon^4\\
&\quad+2\lambda^3(1-\lambda)^3(1-2\lambda)(2-\lambda+\lambda^2)\sum a_1(b_1-c_1)^2\epsilon^6\\
&\quad+2\lambda^3(1-\lambda)^3(1-2\lambda)(2-3\lambda+3\lambda^2)\sum a_1\sum(b_1-c_1)^2\epsilon^6\\
&\quad+8\lambda^4(1-\lambda)^4(1-\lambda+\lambda^2)\sum(b_1-c_1)(b_2-c_2)\epsilon^6+\cdots
\end{aligned}
$$

$$D=\lambda^6(1-\lambda)^6(b_1-c_1)^2(c_1-a_1)^2(a_1-b_1)^2\epsilon^{12}+\cdots$$

in which the summations are symmetrizations in a, b, c. On the other hand, we need the following classical identities:

$$\theta_{10}{}^4+\theta_{01}{}^4=\theta_{00}{}^4 \qquad (\theta_{00}\theta_{10}\theta_{01})^8=2^8\delta$$

$$\frac{d}{d\omega}\log(\theta_{10}/\theta_{00})=\frac{\pi i}{4}\theta_{01}{}^4 \qquad \frac{d}{d\omega}\log(\theta_{00}/\theta_{01})=\frac{\pi i}{4}\theta_{10}{}^4.$$

Using some of them, we get

$$\left.\begin{matrix}a_1\\ b_1\\ c_1\end{matrix}\right\}=\pi i\lambda(1-\lambda)\theta_{00}(\tau_1)^4\frac{d}{d\tau_2}\left\{\begin{matrix}\log\theta_{00}(\tau_2)\theta_{10}(\tau_2)\\ \log\theta_{10}(\tau_2)\theta_{01}(\tau_2)\\ \log\theta_{01}(\tau_2)\theta_{00}(\tau_2).\end{matrix}\right.$$

Therefore, if we put

$$t=\delta(\tau_1)\delta(\tau_2)(\pi\epsilon)^{12},$$

we have

$$D/A^5=2^{-1}3^{-5}t+\cdots.$$

Furthermore, since the Weierstrass invariant j and λ are related as

$$j=2^8(1-\lambda+\lambda^2)^3/\lambda^2(1-\lambda)^2,$$

we have

$$(B/A^2)^3=2^{-12}3^{-6}j(\tau_1)j(\tau_2)t+\cdots.$$

Finally, powers of C/A^3 themselves are of no use, because $3C/A^3$ and B/A^2 have the same leading term. However, if we examine their difference, terms involving a_2, b_2, c_2 cancel each other and we get

$$((3C-AB)/A^3)^2 = 2^{-12}3^{-6}(j(\tau_1)-2^6 3^3)(j(\tau_2)-2^6 3^3)t+\cdots.$$

3. We, now, introduce following set of uniformizing parameters

$$x_1 = 2^4 3^2 B/A^2 \qquad x_2 = 2^6 3^3(3C-AB)/A^3 \qquad x_3 = 2\cdot 3^5 D/A^5$$

of $X = \operatorname{proj} \boldsymbol{C}[A,B,C,D]$ at the point $B=C=D=0$. Also, in view of the calculation we made in the previous section, we put

$$y_1 = x_1^3/x_3 \qquad y_2 = x_2^2/x_3.$$

Then, the normalization of the integral ring $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2]$ (in its field of fractions) is generated by

$$y_3 = x_1^2 x_2/x_3.$$

The proof is as follows. Clearly y_3 is an element of the field of fractions. Moreover, since we have $y_3^2 = x_1 y_1 y_2$, certainly y_3 is integral over the ring. We have only to show, therefore, that $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2, y_3]$ is normal. However, if we operate a cyclic group of order six to a ring of polynomials in three letters t_1, t_2, t_3 as

$$t_1, t_2, t_3 \to \zeta^2 t_1, \zeta^3 t_2, \zeta^5 t_3 \quad (\zeta^6 = 1),$$

the ring of invariant elements of $\boldsymbol{C}[t_1, t_2, t_3]$ can be identified with $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2, y_3]$, in fact, in the following way

$$x_1 = t_1 t_3^2 \qquad x_2 = t_2 t_3^3 \qquad x_3 = t_3^6$$
$$y_1 = t_1^3 \qquad y_2 = t_2^2 \qquad y_3 = t_1^2 t_2 t_2.$$

Therefore, certainly $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2, y_3]$ is normal, and this proves the assertion. Also, we can write down very easily a base of the ideal of relations of x_1, x_2, x_3, y_1, y_2, y_3 consisting of six polynomials. We also note that, if V_0 and V are the affine varieties with $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2]$ and $\boldsymbol{C}[x_1, x_2, x_3, y_1, y_2, y_3]$ as their co-ordinate rings, the holomorphic map of V to V_0 associated with the inclusion of the rings is a bijection. In fact, over the point of V_0 not on

$x_3 = 0$, there exists only one point, and over the point of V_0 on $x_3 = 0$, there also exists only one point, because we necessarily have $y_3 = 0$.

We shall, now, show that the variety of moduli and the variety V form an open affine covering of $F_2 = \Gamma_2 \backslash \mathfrak{S}_2$. We recall that the variety of moduli is the complement of $D = 0$ in X and it is isomorphic to a quotient variety of a three dimensional affine space modulo a cyclic group of order five operating in this space as follows

$$t_j \to \zeta^j t_j \ . \ j = 1, 2, 3 \quad (\zeta^5 = 1).$$

Since the three lambdas considered as meromorphic functions in $\mathfrak{S}_2$ by means of the expressions in terms of theta-functions of zero argument undergo seven-hundred twenty transformations by operations of Γ_2, which are just automorphisms of the field of lambdas relative to the field of absolute invariants, the absolute invariants can be considered as meromorphic functions in $\mathfrak{S}_2$ invariant by operations of Γ_2. Therefore, the absolute invariants are meromorphic functions on the the corresponding quotient variety F_2. We shall show that, if we denote by $(x_3) = (x_3)_0 - (x_3)_\infty$ the divisor on F_2 of the absolute invariant x_3, elements of the co-ordinate ring of the variety of moduli are holomorphic in F_2 minus the support of $(x_3)_0$, simply $F_2 - \text{supp.}(x_3)_0$, and elements of the co-ordinate ring of V are holomorphic in $F_2 - \text{supp.}(x_3)_\infty$. First of all, if a point τ of $\mathfrak{S}_2$ corresponds to a curve of genus two, values of the three lambdas at τ are different from each other and from 0, 1, ∞. Therefore, the A, B, C, D, written as polynomials in the lambdas, are all finite at τ and D is different from 0. On the other hand, if τ corresponds to a product of elliptic curves, i. e., if τ is (equivalent to) a point on $\epsilon = 0$, we have $x_1 = x_2 = x_3 = y_3 = 0$ and

$$y_1 = j(\tau_1) j(\tau_2) \qquad y_2 = (j(\tau_1) - 2^6 3^3)(j(\tau_2) - 2^6 3^3).$$

Therefore, $\text{supp.}(x_3)_0$ consists of points of F_2 representing products of elliptic curves and it has no point in common with $\text{supp.}(x_3)_\infty$. Also, elements of the co-ordinate ring of the variety of moduli are holomorphic in $\mathfrak{S}_2$ at points τ which correspond to curves of genus two, hence they are holomorphic in $F_2 - \text{supp.}(x_3)_0$. Furthermore, the corresponding holomorphic map is a bijection of $F_2 - \text{supp.}(x_3)_0$ to the variety of moduli. On the other hand, the six absolute invariants x_1, x_2, x_3, y_1, y_2, y_3 are all holomorphic in $F_2 - \text{supp.}(x_3)_\infty$. It is clear that they are holomorphic in $F_2 - \text{supp.}(x_3)$. However, their expansions in ϵ show immediately that they are also holomorphic at every point of $\text{supp.}(x_3)_0$. Furthermore, the corresponding holomorphic map is a bijection of $F_2 - \text{supp.}(x_3)_\infty$ to V. In fact, it gives a bijection of $F_2 - \text{supp.}(x_3)$ to V

minus $x_3=0$ and of supp.$(x_3)_0$ to the set of points of V on $x_3=0$. Now, since the variety of moduli and V are both normal, the bijective holomorphic maps are necessarily isomorphisms. In the present case, since the varieties concerned are "V-manifolds" [cf. 8], we have only to use an elementary lemma on removable singularities [2]. This completes the proof.

The observations made so far permit us also to prove the following important lemma:

LEMMA. *The field of meromorphic functions on F_2 can be identified with the field of absolute invariants such that holomorphic functions correspond to constants.*

Since this is an immediate consequence of a general theorem proved by Baily [1] (and of some properties of absolute invariants), we shall give only an outline. Knowing the absolute invariants which generate the co-ordinate rings of the variety of moduli and V, we can easily find a set of projective invariants of the same degree which gives a projective embedding of F_2. In this way, we get a normal projective variety which is a compactification of F_2 such that the complement is one dimensional. This is all we need. Actually, the compactification is a "V-manifold" and the elementary lemma on removable singularities is again sufficient.

III. Ring of algebraic modular forms.

4. In general, if

$$M=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is an element of $Sp(n,\boldsymbol{R})$, the jacobian of the corresponding transformation in $\mathfrak{S}_n$

$$\tau\rightarrow M\tau=(a\tau+b)(c\tau+d)^{-1}$$

is known to be $\det(c\tau+d)^{-n-1}$. We need this fact only for elements of $Sp(n,\boldsymbol{Z})$ (in the case $n=2$) and an easy proof is to verify it for some generators of $Sp(n,\boldsymbol{Z})$. This being remarked, a modular form ψ of degree n is defined by the following two conditions:

(1) *For every element M of $Sp(n,\boldsymbol{Z})$, ψ satisfies a functional equation of the form*

$$\psi(M\tau)=\det(c\tau+d)^{w}\psi(\tau)$$

with some even integer w;

(2) *it is holomorphic in* $\mathfrak{S}_n$.

The integer w in (1) is called the weight of the modular form. In the above definition, the elliptic case $n=1$ is exceptional and we have to assume that ψ is holomorphic also at $i\infty$. More precisely, since, in general, a modular form is invariant under a transformation of the form $\tau \to \tau + b$, in which b is a symmetric integer matrix of degree n, it admits a Fourier expansion of the form

$$\sum_T a(T) e(tr(T\tau)).$$

The summation extends over all symmetric half-integer matrices (i.e. symmetric matrices such that diagonal coefficients and twice all other coefficients are integers) of degree n. Since the modular form is also invariant under a transformation of the form $\tau \to u\tau{}^t u$, in which u is a unimodular integer matrix of degree n, the coefficient $a(T)$ depends only on the class of T in the sense $a({}^t uTu) = a(T)$. Now, a closer examination first made by Koecher shows that $a(T)$ is zero unless T is positive semi-definite [5]. The elliptic case is, however, exceptional and this is not a consequence but an assumption. At any rate, we shall not use this theorem of Koecher and, in the case $n=2$, it will simply come out from our later considerations.

If we consider the set of finite sums of modular forms, we get a subring of the ring of all holomorphic functions in $\mathfrak{S}_n$ and it is obviously graded. We are interested in the structure of this graded ring in the simplest unknown case $n=2$. We say that a modular form (in the case $n=2$) is algebraic if its weight is a non-negative integer divisible by six. Finite sums of algebraic modular forms form a graded subring of the graded ring of all modular forms and we shall determine its structure.

We take an algebraic modular form ψ of weight w and, using the absolute invariants x_1, x_2, x_3, we consider

$$\phi = \psi(\tau)\,(\partial(x_1, x_2, x_3)/\partial(\tau_1, \tau_2, \tau_{12}))^{-w/3}.$$

Then, it is a meromorphic function in $\mathfrak{S}_2$ and, because of the property of the jacobian, it is invariant under operations of Γ_2; hence it is meromorphic on F_2. The lemma in Section 3 implies, therefore, that ϕ is a rational function of x_1, x_2, x_3. Suppose, conversely, that ϕ is a rational function of x_1, x_2, x_3 and consider

$$\psi(\tau) = \phi\,(\partial(x_1, x_2, x_3)/\partial(\tau_1, \tau_2, \tau_{12}))^{w/3}.$$

Then, it satisfies the functional equation, but it may not have the property

(2). We shall, therefore, try to obtain necessary and sufficient conditions (in terms of the rational function ϕ) for ψ to be holomorphic in $\mathfrak{S}_2$.

We first recall that the three lambdas are holomorphic in $\mathfrak{S}_2$ minus those points which are equivalent to points on $\epsilon = 0$. Therefore, we get a holomorphic map of this open subvariety of $\mathfrak{S}_2$ to the space of lambdas, which is a three dimensional affine space minus nine planes defined by $D = 0$, and this map is surjective. The point is that we have an unramified covering, i. e. the holomorphic map is a local isomorphism. In fact, the inverse map is given locally by representing τ as a part of a period matrix belonging to a Rosenhain normal form, and it is certainly holomorphic. We can construct another proof using the fact that no operation of Γ_2 belonging to the so-called principal congruence group modulo 2 has a fixed point in the said open subvariety of $\mathfrak{S}_2$. On the other hand, the lambda space is a ramified covering of the variety of moduli and the ramification takes place along those points which represent curves of genus two having "many automorphisms," i. e. at the singular point $A = B = C = 0$ and along a surface which corresponds to lambda triples satisfying $\lambda_3 - \lambda_1 + \lambda_1\lambda_2 - \lambda_3\lambda_1 = 0$. For this, the reader is referred to the last section of AVM. This surface is related to the "skew-invariant" in the following way. Using the same notation as in Section 1, we consider the following expression:

$$u_0^{15} \prod_{\text{fifteen}} \det \begin{bmatrix} 1 & \xi_1 + \xi_2 & \xi_1\xi_2 \\ 1 & \xi_3 + \xi_4 & \xi_3\xi_4 \\ 1 & \xi_5 + \xi_6 & \xi_5\xi_6 \end{bmatrix}.$$

Then, since each factor has an invariant property, the symmetrized product defines a projective invariant E of degree fifteen. Since the degree is odd, it can not possibly be expressed by A, B, C, D, but its square E^2 can be expressed by these projective invariants. This being remarked, if we express E in terms of the lambdas, we get fifteen distinct irreducible factors one of which is $\lambda_3 - \lambda_1 + \lambda_1\lambda_2 - \lambda_3\lambda_1$. Therefore, the surface in question is defined by $E^2 = 0$ on the variety of moduli and the ramification index is two. We need also an expansion of E^2 into power-series of ϵ. The calculation is straightforward and we get

$$E^2/A^{15} = 2^{-28}3^{-15}(j(\tau_1) - j(\tau_2))^2 t^2 + \cdots.$$

5. Now, consider $\mathfrak{S}_2$ minus points equivalent to points on $\epsilon = 0$ and

points on $A=0$. This open subvariety is ramified over the variety of moduli minus points on $A=0$ along $E^2=0$. We shall translate the condition for ψ to be holomorphic in the said open subvariety of $\mathfrak{S}_2$. We observe that the co-ordinate ring of the variety of moduli minus points on $A=0$ is generated over $\boldsymbol{C}$ by absolute invariants whose denominators are power-products of A and D. Therefore, the co-ordinate ring in question is the ring of fractions of $\boldsymbol{C}[x_1, x_2, x_3]$ with respect to powers of x_3. Since the ramification index is two along $E^2=0$, the condition we are looking for is that ϕ multiplied by E^2/A^{15} to the power w over 6 is in the co-ordinate ring, i. e. that ϕ is a linear combination of

$$(A^{15}/E^2)^{w/6} x_1^{e_1} x_2^{e_2} / x_3^{e_3}$$

in which e_1, e_2, e_3 are integers and e_1, e_2 are non-negative. We shall next write down the condition for ψ to be holomorphic at points on $A=0$. Since the condition is (by the lemma on removable singularities) along $A=0$ and not at special points on $A=0$, we have only to examine those points of it which are not equivalent to points on $\epsilon=0$ and not on $BE^2=0$. Then, instead of the τ's, we can use

$$u_1 = x_1^2/x_3 = 2^7 3^{-1} AB^2/D \qquad u_2 = x_1^5/x_3^2 = 2^{18} B^5/D^2$$
$$u_3 = x_1 x_2/x_3 = 2^9 B(3C - AB)/D$$

as local co-ordinates. Since we have

$$x_1 = u_2/u_1^2 \qquad x_2 = u_2 u_3/u_1^3 \qquad x_3 = u_2^2/u_1^5,$$

we get

$$(A^{15}/E^2)^{w/6} (x_1^{e_1} x_2^{e_2}/x_3^{e_3}) (\partial(x_1, x_2, x_3)/\partial(u_1, u_2, u_3))^{w/3}$$
$$= {}^{*}u_1^{-2e_1-3e_2+5e_3-7w/6}\, u_2^{e_1+e_2-2e_3+w} u_3^{e_2}$$

in which * is a unit depending on w but not on e_1, e_2, e_3. Therefore, for a fixed w, the condition is that a linear combination of the form

$$\sum \text{const.}\, u_1^{-2e_1-3e_2+5e_3-7w/6}\, u_2^{e_1+e_2-2e_3+w} u_3^{e_2}$$

be finite along $u_1=0$. However, in case two distinct triples (e) and (e') give the same exponent of u_1 in the above linear combination, certainly they give different exponents either to u_2 or to u_3. Consequently, for each (e), the exponent of u_1 has to be non-negative. Therefore, the condition is simply

(C1) $$5e_3 - 7w/6 \geqq 2e_1 + 3e_2$$

for every $(e) = (e_1, e_2, e_3)$. In particular, the integer e_3 is also non-negative. Finally, we shall write down the condition for ψ to be holomorphic at those points which are equivalent to points on $\epsilon = 0$, i.e. along $\epsilon = 0$. In order to get this condition, we shall use the parameters t_1, t_2, t_3 introduced in Section 3. Since t_3 is of order two in ϵ, however, we have to use u defined by $t_3 = u^2$. Then, we have

$$(A^{15}/E^2)^{w/6}(x_1{}^{e_1}x_2{}^{e_2}/x_3{}^{e_3})\,(\partial(x_1, x_2, x_3)/\partial(t_1, t_2, u))^{w/3}$$
$$= {}^{*}\, t_1{}^{e_1}t_2{}^{e_2}u^{2(2e_1+3e_2-6e_3)+3w}$$

in which * is a unit depending on w but not on e_1, e_2, e_3. Therefore, in the same way as before, we get

$$\text{(C2)} \qquad 2e_1 + 3e_2 \geqq 6e_3 - \tfrac{3}{2}w$$

for every $(e) = (e_1, e_2, e_3)$. We have thus finished the translation of the condition completely. The result can, obviously, be stated by saying that there exists a monomorphism of the graded ring of algebraic modular forms to the graded ring of multi-canonical differentials on X and that the image ring consists of finite sums of

$$(A^{15}/E^2)^{w/6}(x_1{}^{e_1}x_2{}^{e_2}/x_3{}^{e_3})\,(dx_1dx_2dx_3)^{w/3}$$

in which e_1, e_2, e_3 are non-negative integers satisfying both (C1) and (C2). We note that these differentials are linearly independent and, for a given w, we can compute the number of such differentials quite easily. For instance, we get 1, 1, 3, 4, 8, 11, $\cdots$ for $w = 0, 6, 12, 18, 24, 30, \cdots$.

6. We shall examine the structure of the graded ring of algebraic modular forms. For this purpose, we put $w = 6e_4$ and consider an additive monoid of non-negative integer quadraples (e_1, e_2, e_3, e_4) satisfying

$$5e_3 - 7e_4 \geqq 2e_1 + 3e_2 \geqq 6e_3 - 9e_4.$$

The connection between the ring in question and this monoid is clear. We shall determine the structure of this monoid. If we replace the above two inequalities by equalities, we get a submonoid, and this submonoid is generated by $(0,1,2,1)$ and $(3,0,4,2)$. Now, suppose that (e_1, e_2, e_3, e_4) is an arbitrary element of the monoid. Then, the difference $(e_1, e_2, e_3, e_4) - e_2(0,1,2,1) = (e_1, 0, e_3 - 2e_2, e_4 - e_2)$ is again an element of the monoid, and this reduces our consideration to its submonoid consisting of elements of that type. In fact, if four real numbers $\epsilon_1, \epsilon_2 = 0, \epsilon_3, \epsilon_4$ satisfy the above inequalities and if ϵ_1 is non-negative, the others are also non-negative, This simple remark will be used repeatedly. Suppose, now, that $(e_1, 0, e_3, e_4)$ is an arbitrary element of the submonoid. This time we subtract $(3, 0, 4, 2)$ from $(e_1, 0, e_3, e_4)$ so that the first coefficient of the difference becomes 0, 1 or 2. Then, as we know, the other coefficients are non-negative; hence the difference is still an element of the monoid. Suppose, therefore, that $(e_1, 0, e_3, e_4)$ is an element of the monoid and that $e_1 = 0, 1, 2$. If we have $e_1 = 0$, the element is in the submonoid consisting of elements of the form $(0, 0, e_3, e_4)$ in which e_3, e_4 are, of course, non-negative and satisfy

$$5e_3 - 7e_4 \geqq 0 \qquad 3e_4 - 2e_3 \geqq 0.$$

If we replace one of the inequalities by an equality, we get $(0, 0, 7, 5)$ and $(0, 0, 3, 2)$, and these elements generate the submonoid. In fact, the element $(0, 0, e_3, e_4)$ decomposes into $5e_3 - 7e_4$ times $(0, 0, 3, 2)$ and $3e_4 - 2e_3$ times $(0, 0, 7, 5)$. If we have $e_1 = 1$, we try to reduce the element $(1, 0, e_3, e_4)$ using $(0, 0, 3, 2)$ and $(0, 0, 7, 5)$. The reduction will fail to work if both $(1, 0, e_3, e_4) - (0, 0, 3, 2)$ and $(1, 0, e_3, e_4) - (0, 0, 7, 5)$ are outside the monoid. This happens if and only if we have both $5e_3 - 7e_4 = 2$ and $2e_3 - 3e_4 = 0$. Thus, we get $(1, 0, 6, 4)$. Finally, if we have $e_1 = 2$, we again try to reduce the element $(2, 0, e_3, e_4)$ using $(0, 0, 3, 2)$ and $(0, 0, 7, 5)$. The reduction fails to work if and only if we have $5e_3 - 7e_4 = 4$ and $2e_3 - 3e_4 = 1$. This gives $(2, 0, 5, 3)$. We have, thus, shown that our monoid is generated by the following six elements:

$$\begin{array}{ccc} (0,1,2,1) & (0,0,3,2) & (3,0,4,2) \\ (0,0,7,5) & (2,0,5,3) & (1,0,6,4). \end{array}$$

In other words, we have found generators of the graded ring of algebraic modular forms. The structure of this ring will be known, therefore, if we determine all possible relations between the above six generators. For this purpose, we observe that a column vector with six coefficients is annihilated by the matrix

$$\begin{pmatrix} 0 & 0 & 3 & 0 & 2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 3 & 4 & 7 & 5 & 6 \\ 1 & 2 & 2 & 5 & 3 & 4 \end{pmatrix}$$

if and only if the coefficients are of the form $0, 0, x, y, -2x+y, x-2y$. In particular, we have

$$2(2,0,5,3)=(3,0,4,2)+(1,0,6,4)$$
$$2(1,0,6,4)=(0,0,7,5)+(2,0,5,3)$$
$$(2,0,5,3)+(1,0,6,4)=(3,0,4,2)+(0,0,7,5).$$

Using these relations, we can lower the sum of multiplicities of $(2,0,5,3)$ and $(1,0,6,4)$ in any expression of an element of the monoid by the six generators as long as it is at least equal to two. In this way, every element of the monoid can be expressed in one of the three forms

$$p(0,1,2,1)+q(0,0,3,2)+r(3,0,4,2)+s(0,0,7,5)$$
$$+\,0,\ (2,0,5,3),\ (1,0,6,4)$$

in which p, q, r, s are non-negative integers. Moreover, the expression is unique. Thus, the structure of the ring of algebraic modular forms is completely determined. In particular, the dimension N_w of the complex vector space of modular forms of weight w is, in case w is (non-negative and) of the form $6m$, equal to the number of partitions of m into the form $p+2(q+r)+5s+0, 3, 4$. There are ten formulas for N_w according to the values of w modulo sixty and, for example, we have

$$N_{60k}=150\binom{k}{3}+190\binom{k}{2}+51\binom{k}{1}+1.$$

IV. Main theorems.

7. Since the structure of the ring of algebraic modular forms is determined, using this result, we shall start investigating the ring of all modular forms in connection with Eisenstein series. In general, if τ is a point of $\mathfrak{S}_n$, the Eisenstein series of degree n and of weight w is defined as follows

$$\psi_w(\tau)=\sum_{\{c,d\}}\det(c\tau+d)^{-w}.$$

The summation extends over all classes of coprime symmetric pairs, i. e. over all inequivalent bottom rows of elements of $Sp(n,\mathbf{Z})$ with respect to left multiplications by unimodular integer matrices of degree n. It is a classical theorem of H. Braun that the series is absolutely convergent for $w>n+1$ and represents a modular form of weight w. The modular forms defined by

Eisenstein series are connected with positive-definite quadratic forms by their Fourier expansions. In the Einführung, Siegel gives a formula for the coefficients of Fourier expansions of Eisenstein series in terms of what he called p-adic densities [9]. Therefore, in order to calculate the Fourier coefficients using this formula, we have to calculate certain p-adic densities for all p including $p=2$. In some cases, however, the so-called Siegel main theorem can be used to go around this tedious calculation. On the other hand, if $a(T)$ is a Fourier coefficient of Eisenstein series of degree n, the said Siegel formula shows that

$$a(T_1)=a\begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}$$

is a Fourier coefficient of Eisenstein series of degree $n-1$ and of the same weight. This allows us, in some cases, to reduce the calculation of the Fourier coefficient from degree n to degree $n-1$, thus finally to the elliptic case. In this case, if we denote the Eisenstein series of weight w by ϕ_w, we have

$$\phi_w(\omega)=1+((2\pi i)^w/\Gamma(w)\zeta(w))\sum_{t=1}^{\infty}(\sum_{d\mid t}d^{w-1})e(t\omega).$$

Also, in the case $n=2$, we have the following table:

$$w=4$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}=2^4 3\cdot 5 \qquad a\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}=2^4 3^3 5$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}=2^5 3^3 5\cdot 7 \qquad a\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}=2^7 3\cdot 5\cdot 7$$

$$w=6$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}=-2^3 3^2 7 \qquad a\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}=-2^3 3^3 7\cdot 11$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}=2^4 3^3 5\cdot 7\cdot 11 \qquad a\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}=2^6 3^2 7\cdot 11$$

$$w=8$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}=2^5 3\cdot 5 \qquad a\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}=2^5 3^2 5\cdot 43$$

$$a\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}=2^6 3^2 5\cdot 61 \qquad a\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}=2^8 3\cdot 5\cdot 7$$

$$w = 10$$

$$a\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} = -2^3 3\cdot 11 \qquad a\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix} = -2^3 3^4 11\cdot 19$$

$$a\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = 2^4 3^4 5\cdot 7\cdot 11\cdot 19\cdot 277\cdot 43867^{-1}$$

$$a\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix} = 2^6 3\cdot 7\cdot 11\cdot 19\cdot 809\cdot 43867^{-1}$$

$$w = 12$$

$$a\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} = 2^4 3^2 5\cdot 7\cdot 13\cdot 691^{-1} \qquad a\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix} = 2^4 3^3 5\cdot 7\cdot 13\cdot 683\cdot 691^{-1}$$

$$a\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = 2^5 3^3 5\cdot 7\cdot 13\cdot 19\cdot 23\cdot 2659\cdot 131^{-1} 593^{-1} 691^{-1}$$

$$a\begin{pmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{pmatrix} = 2^7 3^2 5\cdot 7\cdot 13\cdot 23\cdot 1847\cdot 131^{-1} 593^{-1} 691^{-1}.$$

Some of the large prime numbers we find in this table come from Bernoullian numbers. For instance 43867 is the numerator of the ninth Bernoullian number. At any rate, if we observe that we always have $a(0) = 1$, we see that the above table gives Fourier coefficients $a(T)$ of Eisenstein series of weight at most twelve for all (symmetric half-integer positive semi-definite) matrices T satisfying $tr(T) \leqq 2$. Also, in order to avoid a possible misunderstanding by some reader, we note that the table is not really necessary until we start proving explicit rational expressions of absolute invariants by Eisenstein series.

Now, it is known (in the case $n = 2$) that the dimension N_w is zero for $w = 2$ and one for $w = 4, 6, 8$. This fact was proved by Maass [6] and partly by Witt [12]. One of their main ideas is to use a remarkable lemma by Siegel on a unique determination of modular forms (of a given weight) by their first few Fourier coefficients. Incidentally, this lemma shows also that the Fourier coefficients of the five Eisenstein series in the previous table determine these modular forms uniquely. When we started working on the subject, beside some general results, that was practically all we knew about the ring of modular forms. At any rate, we shall not make any use of the above results and they will come out from our subsequent considerations.

We shall denote by $\psi_{e_1e_2e_3e_4}$ the algebraic modular form defined by

$$\psi_{e_1e_2e_3e_4}(\tau) = (2^{-38}3^{-15}A^{15}/E^2)^{e_4}(x_1^{e_1}x_2^{e_2}/x_3^{e_3})\,(\partial(x_1,x_2,x_3)/\partial(\tau_1,\tau_2,\tau_{12}))^{2e_4}.$$

It will become necessary to expand this modular form into a Fourier series or, more precisely, into a power-series of $e(\tau_1)$, $e(\tau_2)$ and $\epsilon = \tau_{12}$. For this calculation, we need following expansions

$$j = 2^6 3^3 \phi_4^3 / (\phi_4^3 - \phi_6^2) = e(-\omega) + 2^3 3 \cdot 31 + 2^2 3^3 1823 \cdot e(\omega) + \cdots$$

$$\partial(x_1, x_2, x_3) / \partial(\tau_1, \tau_2, \tau_{12})$$
$$= 2^7 3^3 (j(\tau_1) - j(\tau_2)) j'(\tau_1) j'(\tau_2) (y_3^{11}/y_1^8 y_2^6)(1/\epsilon) + \cdots .$$

This being remarked, we note that ψ_{0032}, ψ_{0242}, ψ_{3042} form a base of the vector space of modular forms of weight twelve. Since ψ_4^3 is a modular form of the same weight, therefore, it is a linear combination of these modular forms. The coefficients of the linear combination can be calculated with the aid of Fourier expansions. If we observe that the three modular forms are linearly independent also on $\epsilon = 0$, we have only to calculate the Fourier expansion of ψ_4 on $\epsilon = 0$, but this is simply $\phi_4(\tau_1)\phi_4(\tau_2)$. In this way, without really using the Fourier coefficients of ψ_4 itself, we get

$$\psi_{3042} = 2^{36} 3^{12} \pi^{12} \psi_4^3 .$$

We can also use the following argument to see that ψ_{3042} and ψ_4^3 differ only by a constant factor. Because of $N_{12} = 3$, we have $N_4 \leqq 1$, hence $N_4 = 1$. On the other hand, the divisor of ψ_{3042} in $\mathfrak{S}_2$ is three times another divisor. Therefore, its cubic root divided by ψ_4 defines a "multiplicative function," i. e. a function whose absolute value is single-valued, on F_2. However, using the structure of the compactification, we can conclude very easily that the function itself is single-valued. Consequently, the cubic root of ψ_{3042} is a modular form of weight four and, because of $N_4 = 1$, it differs from ψ_4 by a constant factor. In this way, we can minimize the use of numerical computation. Besides, the exact constant factor is not necessary in proving most of our theorems.

At any rate, if we adjoin ψ_4 to the ring of algebraic modular forms, we get a new graded ring. We shall show that the normalization, say S, of this graded ring (in its field of fractions) is precisely the graded ring of all modular forms. Suppose that ψ is an arbitrary modular form of weight w. Then, first of all, the weight w is non-negative. Otherwise, the product $\psi^{12}\psi_4^{-3w}$ will be a holomorphic function on F_2, hence it is a constant by a lemma in Section 3. In particular, both ψ and ψ_4 will be units in the ring of holomorphic functions in $\mathfrak{S}_2$. But, certainly ψ_4 is not a unit. Hence w is non-negative. This being remarked, let $2e$ be the least residue of w modulo six. Then, the product $\psi\psi_4^e$ is certainly an algebraic modular form.

Hence, every modular form is in the field of fractions. Furthermore, since the sixth power of any modular form is an algebraic modular form, it is integral over the ring. Therefore, every modular form is in S. On the other hand, we know in general that a normalization of a graded integral ring over a field is itself a graded ring [13]. In particular, our S is a graded ring. Moreover, a homogeneous element ϕ of S can be expressed as the quotient of two homogeneous elements of the ring, and they are modular forms. Therefore, it satisfies the functional equation. Also, since ϕ is integral over the ring of holomorphic functions in $\mathfrak{S}_2$, it is holomorphic there. This shows that ϕ is a modular form, hence S is the graded ring of modular forms. We shall, now, determine the structure of S explicitly. In doing this, we shall not use the second half of the above proof, i. e. the fact that every homogeneous element of S is a modular form. It will come out as a consequence.

We recall that the ring S is the normalization of a graded ring generated over $\boldsymbol{C}$ by

$$\psi_4, \quad \psi_{0121}, \quad \psi_{0032}, \quad \psi_{0075}, \quad \psi_{2053}, \quad \psi_{1064}.$$

These elements, except ψ_{0121} and ψ_{0032}, are related as

$$\psi_{2053}{}^3 = \text{const.}\, \psi_4{}^6\psi_{0075} \qquad \psi_{1064}{}^3 = \text{const.}\, \psi_4{}^3\psi_{0075}{}^2.$$

Therefore, the field of fractions of $\boldsymbol{C}[\psi_4, \psi_{0075}, \psi_{2053}, \psi_{1064}]$ is of degree of transcendency at most two over $\boldsymbol{C}$. Since the field of fractions of the ring of modular forms is of degree of transcendency four over $\boldsymbol{C}$, the degree of transcendency in question is precisely two and the two elements ψ_{0121} and ψ_{0032} are algebraically independent over that field. Therefore, by recalling that a ring of polynomials with coefficients in a normal integral ring is itself normal, we conclude that, if R is the normalization of $\boldsymbol{C}[\psi_4, \psi_{0075}, \psi_{2053}, \psi_{1064}]$ (in its field of fractions), we get S by just adjoining ψ_{0121} and ψ_{0032} to R. Now, if we put $\chi = \psi_{2053}/\psi_4{}^2$, we have

$$\chi^3 = \text{const.}\, \psi_{0075}.$$

Hence χ is in R, and R is also the normalization of $\boldsymbol{C}[\psi_4, \chi, \psi_{1064}]$. On the other hand, we have

$$\psi_{1064}{}^3 = \text{const.}\, (\psi_4\chi^2)^3,$$

hence $\psi_{1064} = \text{const.}\, \psi_4\chi^2$. Since $\boldsymbol{C}[\psi_4, \chi]$ is certainly normal (as a ring of polynomials in two letters with coefficients in $\boldsymbol{C}$), therefore, we get $R = \boldsymbol{C}[\psi_4, \chi]$, hence finally

$$S = \boldsymbol{C}[\psi_4, \chi, \psi_{0121}, \psi_{0032}].$$

We state our result in the following way:

THEOREM 1. *The graded ring of modular forms is generated over* **C** *by four (algebraically independent) modular forms*

$$\psi_4, \quad \psi_{0121}, \quad \chi = \psi_{2058}/\psi_4^2, \quad \psi_{0082}$$

of respective weight four, six, ten and twelve.

COROLLARY. *The dimension N_w of the complex vector space of modular forms of weight w is equal to the number of non-negative integer solutions of the linear Diophantine equation*

$$w = 4p + 6q + 10r + 12s.$$

8. We have shown that the graded ring of modular forms is generated by four modular forms of which one is the Eisenstein series ψ_4. We shall show that other three can also be expressed by Eisenstein series. First of all, since we have $N_6 = 1$, two modular forms ψ_6 and ψ_{0121} differ only by a constant factor. This constant factor can be determined immediately. In order to make the rest of the argument clear, we shall explicitly use Siegel's operator Φ which maps a modular form of degree n to a modular form of degree $n-1$ of the same weight for $n = 1, 2, 3, \cdots$. If ψ is a modular form of degree n and if $a(T)$ is its Fourier coefficient, the Fourier coefficient of $\Phi\psi$ is given by

$$a(T_1) = a\begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In particular, Eisenstein series are mapped by Φ to Eisenstein series. The operator Φ gives rise to a homomorphism of the graded rings of modular forms [10] and it is almost surjective [6]. In the case $n = 2$, the surjectivity is obvious because the graded ring of elliptic modular forms is generated by the Eisenstein series ϕ_4 and ϕ_6. A modular form is called a cusp form if it is in the kernel of Φ. In the case $n = 2$, therefore, there exists one cusp form and only one up to a constant factor of weight ten and of weight twelve. These cusp forms can be obtained in two ways. Since we have $\phi_4\phi_6 = \phi_{10}$, we see that $\psi_4\psi_6 - \psi_{10}$ vanishes along $\epsilon = 0$, hence, certainly, it is a cusp form. Also, since we have

$$3^2 7^2 \phi_4^3 + 2 \cdot 5^3 \phi_6^2 - 691 \phi_{12} = 0,$$

we get a cusp form by replacing phi by psi. We have thus found cusp forms of weight ten and twelve, and they have Fourier expansions of the following form

$$\psi_4\psi_6 - \psi_{10} = -2^{12}3^5 5^2 7\cdot 53\cdot 43867^{-1} e(\tau_1)e(\tau_2)(\pi\epsilon)^2 + \cdots$$

$$3^2 7^2 \psi_4{}^3 + 2\cdot 5^3\psi_6{}^2 - 691\psi_{12}$$
$$= 2^{13}3^7 5^3 7^2 337\cdot 131^{-1}593^{-1} e(\tau_1)e(\tau_2) + \cdots.$$

Therefore, we shall define normalized cusp forms by

$$\chi_{10} = -43867\cdot 2^{-12}3^{-5}5^{-2}7^{-1}53^{-1}(\psi_4\psi_6 - \psi_{10})$$
$$\chi_{12} = 131\cdot 593\cdot 2^{-13}3^{-7}5^{-3}7^{-2}337^{-1}(3^2 7^2\psi_4{}^3 + 2\cdot 5^3\psi_6{}^2 - 691\psi_{12}).$$

On the other hand, the modular forms χ and ψ_{0082} in Theorem 1 are also cusp forms of weight ten and twelve, and they have Fourier expansions of the form

$$\chi = 2^{54}3^{18}\pi^{18} e(\tau_1)e(\tau_2)(\pi\epsilon)^2 + \cdots$$
$$\psi_{0082} = 2^{36}3^{12}\pi^{12} e(\tau_1)e(\tau_2) + \cdots.$$

Therefore, we have the following theorem:

THEOREM 2. *The three modular forms* ψ_{0121}, χ *and* ψ_{0082} *can be expressed by Eisenstein series in the form*

$$\psi_{0121} = 2^{18}3^6\pi^6\psi_6 \qquad \chi = 2^{54}3^{18}\pi^{18}\chi_{10}$$
$$\psi_{0082} = 2^{36}3^{12}\pi^{12}\chi_{12}.$$

COROLLARY. *The graded ring of modular forms is generated over* $\boldsymbol{C}$ *by the Eisenstein series* ψ_4, ψ_6, ψ_{10} *and* ψ_{12}.

Now, we recall that elements of weight zero of the field of fractions of the graded ring of modular forms are called modular functions. By the lemma in Section 3, modular functions are absolute invariants. The converse is also true, i. e. absolute invariants are modular functions. They can be expressed, therefore, rationally in terms of Eisenstein series. Now, there is a problem once proposed by Siegel to get these rational expressions explicitly [9, p. 604]. In order to solve this problem, we have only to write down the three absolute invariants x_1, x_2, x_3 by ψ_4, ψ_6, χ_{10} and χ_{12}. However, using the relation betwen algebraic modular forms and multi-canonical differentials on X, we immediately get the following result:

THEOREM 3. *The three absolute invariants can be expressed by the four modular forms in the form*

$$x_1 = \psi_4\chi_{10}{}^2/\chi_{12}{}^2 \qquad x_2 = \psi_6\chi_{10}{}^3/\chi_{12}{}^3 \qquad x_3 = \chi_{10}{}^6/\chi_{12}{}^5.$$

Thus, the problem raised by Siegel is completely solved.

9. In this last section, we shall investigate the structure of Satake's compactification in the case $n=2$. Because of Theorem 1, the problem is reduced to determining the structure of

$$Y = \operatorname{proj} \boldsymbol{C}[T_2, T_3, T_5, T_6]$$

in which T_2, T_3, T_5, T_6 are indeterminates of degree four, six, ten, twelve and the ring is graded by the total degree. The projective variety Y admits a covering by four open affine subvarieties Y_2, Y_3, Y_5, Y_6 in which Y_j is the complement of $T_j = 0$ for $j = 2, 3, 5, 6$. We shall start by examining these affine varieties.

First, if we operate a cyclic group of order two to a ring of polynomials in three letters t_1, t_2, t_3 as

$$t_1, t_2, t_3 \rightarrow -t_1, -t_2, t_3,$$

the ring of invariant elements of $\boldsymbol{C}[t_1, t_2, t_3]$ can be identified with the co-ordinate ring of Y_2 in the following way:

$$T_3^{e_3} T_5^{e_5} T_6^{e_6} / T_2^{e_2} = t_1^{e_3} t_2^{e_5} t_3^{e_6}.$$

In fact, the condition on one side is $3e_3 + 5e_5 + 6e_6 = 2e_2$ and the condition on the other side is $e_3 + e_5 \equiv 0 \bmod 2$. These two conditions are clearly equivalent. Since the co-ordinate ring of Y_2 is the subring of $\boldsymbol{C}[t_1, t_2, t_3]$ generated by t_1^2, $t_1 t_2$, t_2^2, t_3, the variety Y_2 is isomorphic to a product of a representative cone of a non-degenerate conic and a straight line. The singular locus of Y_2 is, therefore, the locus of the vertex of the cone, which is a straight line defined by $T_3 = T_5 = 0$. As for Y_3, if we operate a cyclic group of order three to a ring of polynomials in t_1, t_2, t_3 as

$$t_1, t_2, t_3 \rightarrow \zeta t_1, \zeta t_2, t_3 \quad (\zeta^3 = 1),$$

the ring of invariant elements of $\boldsymbol{C}[t_1, t_2, t_3]$ can be identified with the co-ordinate ring of Y_3 in the following way:

$$T_2^{e_2} T_5^{e_5} T_6^{e_6} / T_3^{e_3} = t_1^{e_2} t_2^{e_5} t_3^{e_6}.$$

In fact, the two conditions $2e_2 + 5e_5 + 6e_6 = 3e_3$ and $e_2 + e_5 \equiv 0 \bmod 3$ are equivalent. Since the co-ordinate ring of Y_3 is the subring of $\boldsymbol{C}[t_1, t_2, t_3]$ generated by t_1^3, $t_1^2 t_2$, $t_1 t_2^2$, t_2^3, t_3, the variety Y_3 is isomorphic to a product of a representative cone of a cubic space curve and a straight line. The singular locus of Y_3 is, therefore, the locus of the vertex of the cone, which is a straight line defined by $T_2 = T_5 = 0$. We have to examine two more varieties. If we operate a cyclic group of order five to a ring of polynomials in t_1, t_2, t_3 as

$$t_j \to \zeta^j t_j \qquad j = 1, 2, 3 \qquad (\zeta^5 = 1),$$

the ring of invariant elements of $\boldsymbol{C}[t_1, t_2, t_3]$ can be identified with the co-ordinate ring of Y_5 in the following way:

$$T_2^{e_2}T_3^{e_3}T_6^{e_4}/T_5^{e_5} = t_1^{e_1}t_2^{e_2}t_3^{e_3}.$$

Therefore, as we have shown in AVM, the co-ordinate ring of Y_5 is the subring of $\boldsymbol{C}[t_1, t_2, t_3]$ generated by t_1^5, $t_1^3t_2$, $t_1t_2^2$, $t_1^2t_3$, t_2t_3, $t_1t_3^3$, t_2^5, t_3^5 and the variety Y_5 is isomorphic to the variety of moduli. The singular locus of Y_5 is the point defined by $T_2 = T_3 = T_6 = 0$. Finally, if we operate a cyclic group of order six to a ring of polynomials in t_1, t_2, t_3 as

$$t_1, t_2, t_3 \to \zeta^2 t_1, \zeta^3 t_2, \zeta^5 t_3 \quad (\zeta^6 = 1),$$

the ring of invariant elements of $\boldsymbol{C}[t_1, t_2, t_3]$ can be identified with the co-ordinate ring of Y_6 in the following way:

$$T_2^{e_2}T_3^{e_3}T_5^{e_5}/T_6^{e_6} = t_1^{e_1}t_2^{e_2}t_3^{e_3}.$$

Therefore, the variety Y_6 is isomorphic to the affine variety V in Section 3 and its singular locus consists of two straight lines defined by $T_2 = T_5 = 0$ and by $T_3 = T_5 = 0$. Hence, the singular locus of Y itself consists of a point defined by $T_2 = T_3 = T_6 = 0$ and of two projective straight lines defined by $T_2 = T_5 = 0$ and by $T_3 = T_5 = 0$ intersecting at $T_2 = T_3 = T_5 = 0$.

It is, now, easy to determine Betti numbers b_j of the variety Y for all j. Since Y_2 admits a three dimensional affine space as a two-sheeted covering, if a singular chain of Y does not cover the whole Y_2, twice this chain is homotopic to a singular chain on $T_2 = 0$. Therefore, beside $b_0 = b_6 = 1$, we have $b_5 = 0$. Since the intersection of Y_3 and $T_2 = 0$ is a two dimensional affine space, we have $b_4 = 1$ and $b_3 = 0$. Finally, since the intersection of Y_5 and $T_2 = T_3 = 0$ is a straight line, we have $b_2 = 1$ and $b_1 = 0$. Therefore, the projective variety Y has the same Betti numbers as three dimensional complex projective space.

We have thus investigated the structure of Y which is defined abstractly. We shall, now, translate the results into a language of modular forms. For this purpose, we put

$$T_2 = a\psi_4 \qquad T_3 = b\psi_6 \qquad T_5 = c\psi_4\psi_6 + c'\psi_{10}$$
$$T_6 = d\psi_4^3 + d'\psi_6^2 + d''\psi_{12}$$

in which a, b, c, c', d, d', d'' are constants and $abc'd''$ is different from zero. Then, we get a holomorphic map of F_2 to Y so that the variety of moduli and the variety V are mapped isomorphically to Y_5 and Y_6. Since the verification is straightforward, we shall leave it to the reader. Moreover, the complement of the image is a projective straight line isomorphic to proj $\boldsymbol{C}[\phi_4, \phi_6]$. This can be identified with j-line compactified by a point at infinity, i. e. with the union of F_1 and F_0. Therefore Y is isomorphic to Satake's compactification of F_2 and it is a "V-manifold" whose structure we know completely. We note that the singular locus of Y consists of a point representing the jacobian variety of $y^2 = 1 - x^5$ and of two projective straight lines of points representing products of elliptic curves of which at least one factor is $y^2 = 1 - x^3$ or $y^2 = 1 - x^4$ (allowing another factor to degenerate to $y^2 = 1 - x^2$). Equations in terms of the Eisenstein series of these cases are $\psi_4 = \psi_6 = \psi_{12} = 0$, $\psi_4 = \psi_{10} = 0$ and $\psi_6 = \psi_{10} = 0$.

Appendix. The dimension N_w of the complex vector space of modular forms of weight w is equal to the number of non-negative integer solutions of $w = 4p + 6q + 10r + 12s$. This is derived from a structure theorem of the graded ring of modular forms. It is known, on the other hand, that A. Selberg has a general "trace formula" which contains a formula for N_w as a special case. However, since his formula is not given in a "finite form" even in the case of modular group of degree two (because of serious complications coming from non-compact boundaries), we did not try to use this formula. Nevertheless, we feel that we have to spend a few lines connected with this approach to the problem of determining the structure of the graded ring of modular forms, because it is the only general method we can think of at the present moment.

The result we have for N_w can be written in the form

$$N_{2k} = \mathrm{Res}_{x=0}\, dx/(1 - x^2)(1 - x^3)(1 - x^5)(1 - x^6)x^{k1+}$$

for $k = 0, \pm 1, \pm 2, \cdots$. We transform this formula using the theorem of residues and calculate all residues. In this way, we get the following formula

$$\begin{aligned} N_{2k} = {} & 2^{-4}3^{-3}5^{-1}(2k^3 + 48k^2 + 347k + 728) \\ & + (-1)^k 2^{-4}3^{-1}(k + 8) \\ & + 2^{-2}3^{-4}(\rho^{2k+1} + \bar{\rho}^{2k+1})(6k + 41) \\ & + 2^{-1}3^{-4}7(\rho^{2k} + \bar{\rho}^{2k}) \\ & + 2^{-2}3^{-2}(\rho^k + \bar{\rho}^k + \rho^{k+1} + \bar{\rho}^{k+1}) \\ & + 5^{-2}\sum_{\zeta^5=1} \zeta^k - \zeta^{k+1} \end{aligned}$$

for $k = 0, 1, 2, \cdots$, where ρ stands for $e(\frac{1}{8})$ and $\bar{\rho}$ is its complex conjugate. Now, it may be possible to prove this formula by Selberg's method. Therefore, it will be of some importance if we shall show that this dimension formula implies the structure theorem. At any rate, we get the two cusp forms χ_{10} and χ_{12} as in Section 8. We shall show that ψ_4, ψ_6, χ_{10} and χ_{12} are algebraically independent over $\boldsymbol{C}$. If they are not algebraically independent, there exists a homogeneous element $P(T)$ not equal to zero in the graded ring $\boldsymbol{C}[T_2, T_3, T_5, T_6]$ (of Section 9) satisfying $P(\psi_4, \psi_6, \chi_{10}, \chi_{12}) = 0$. We take as $P(T)$ one which has a smallest degree and write it in the form $P_0(T_2, T_3, T_5, T_6)T_5 + P_1(T_2, T_3, T_6)$. Now, we introduce an operator Ψ to be the restriction of (finite sums of) modular forms to the subvariety of $\mathfrak{S}_2$ defined by $\epsilon = 0$. Then Ψ is a homomorphism and χ_{10} is in the kernel. Therefore, applying Ψ to

$$P_0(\psi_4, \psi_6, \chi_{10}, \chi_{12})\chi_{10} + P_1(\psi_4, \psi_6, \chi_{12}) = 0,$$

we get $P_1(\Psi\psi_4, \Psi\psi_6, \Psi\chi_{12}) = 0$. We shall show that $\Psi\psi_4$, $\Psi\psi_6$ and $\Psi\chi_{12}$ are algebraically independent (over $\boldsymbol{C}$). For this purpose, we observe that we have a relation of the form

$$\begin{aligned}\Psi\chi_{12}(\tau_1, \tau_2) = {} & \text{const.}\, \phi_4(\tau_1)^3\phi_4(\tau_2)^3 + \text{const.}\, \phi_6(\tau_1)^2\phi_6(\tau_2)^2 \\ & + \text{const.}\, (\phi_4(\tau_1)^3\phi_6(\tau_2)^2 + \phi_4(\tau_2)^3\phi_6(\tau_1)^2)\end{aligned}$$

in which the third constant coefficient is different from zero. On the other hand, if x, y, x', y' are algebraically independent, certainly xx', yy', $x^3y'^2 + x'^3y^2$ are also algebraically independent. Therefore $\Psi\psi_4$, $\Psi\psi_6$, $\Psi\chi_{12}$ are algebraically independent. Hence, we have $P_1(T_2, T_3, T_6) = 0$. This will imply that $P_0(T)$ is different from zero and $P_0(\psi_4, \psi_6, \chi_{10}, \chi_{12}) = 0$. Since $P_0(T)$ is of a smaller degree than $P(T)$, we get a contradiction. Therefore ψ_4, ψ_6, χ_{10} and χ_{12} are algebraically independent. Then, by the dimension formula, the graded ring generated by these four modular forms will be the ring of all modular forms. This is the structure theorem.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] W. L. Baily, "Satake's compactification of V_n," *American Journal of Mathematics*, vol. 80 (1958), pp. 348-364.

[2] S. Bochner-W. T. Martin, *Several complex variables*, Princeton (1948).

[3] Séminaire H. Cartan, *Fonctions automorphes*, E. N. S. (1957-58).

[4] J. Igusa, "Arithmetic variety of moduli for genus two," *Annals of Mathematics*, vol. 72 (1960), pp. 612-649 (cited AVM).

[5] M. Koecher, "Zur Theorie der Modulfunktionen n-ten Grades, I," *Mathematische Zeitschrift*, vol. 59 (1954), pp. 399-416.

[6] H. Maass, "Über die Darstellung der Modulformen n-ten Grades durch Poincarésche Reihen," *Mathematische Annalen*, vol. 123 (1951), pp. 125-151.

[7] G. Rosenhain, *Abhandlung über die Functionen zweier Variabler mit vier Perioden*, 1851, Ostwald's Klassiker der Exacten Wissenschaften, no. 65 (1895).

[8] I. Satake, "On the compactification of the Siegel space," *Journal of the Indian Mathematical Society*, vol. 20 (1956), pp. 259-281.

[9] C. L. Siegel, "Über die analytische Theorie der quadratischen Formen," *Annals of Mathematics*, vol. 36 (1935), pp. 527-606.

[10] ———, "Einführung in die Theorie der Modulfunktionen n-ten Grades," *Mathematische Annalen*, vol. 116 (1939), pp. 617-657.

[11] A. Weil, "Zum Beweis des Torellischen Satzes," *Göttingen Nachrichten*, Nr. 2 (1957), pp. 33-53.

[12] E. Witt, "Eine Identität zwischen Modulformen zweiten Grades," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 14 (1941), pp. 323-337.

[13] O. Zariski, "Some results in the arithmetic theory of algebraic varieties," *American Journal of Mathematics*, vol. 61 (1939), pp. 249-294.

A CHARACTERIZATION OF THE SIMPLE GROUPS $SL(2, 2^a)$: CORRECTION.*

By WALTER FEIT.

In a previous paper [10] a condition was stated which characterized the simple groups $SL(2, 2^a)$. Professor M. Suzuki has pointed out that the proof of Lemma 6.9 is incorrect, the elements $z_i x_1$ appearing there should be replaced by $x_i z_4$. This invalidates the argument of Section 6 in [10]. The purpose of this note is to replace Section 6 in [10] by a new Section 6. The notation and bibliography are exactly as in [10] except for the additional references at the end of this note.

6. Completion of the proof of Case II. Throughout this section it will be assumed that the group G satisfies the hypotheses stated at the beginning of Section 3. Furthermore, there exists an element y in $C(S_2) - \{1\}$ with the property that $C(y)$ is not contained in $N(S_2)$. Thus the results of Sections 3 and 5 can be used. Our aim is to reach a contradiction. Once this has been achieved it will, together with Lemma 4.4 complete the proof of Theorem 1 in all cases.

LEMMA 6.1. *D is a cyclic p-group for some prime p and d divide* $(2^a + 1)$.

Proof. By Lemma 2.1 L contains a cyclic subgroup K of order $(2^a + 1)$. Let K_p, D_p denote a Sylow p-group of K, D respectively. Let $K_p{}^0$, $D_p{}^0$ denote the subgroup of K_p, D_p respectively generated by all elements of order p.

If D is not a p-group then ([7] page 175) it is possible to choose primes $p \neq q$ such that $D \subset N(D_q)$ and $D_p \cap D' = \{1\}$. Furthermore D_p, D_q are cyclic and $D \subset C(D_p{}^0)$. By Lemmas 5.1 and 5.2 this implies that $C(K_p{}^0 \times D_p{}^0) \subset L \times D$. Therefore D_q is a characteristic subgroup of $C(K_p{}^0 \times D_p{}^0)$. Hence $N(K_p{}^0 \times D_p{}^0) \subset N(D_q) \subset L \times D$ by Lemma 5.2. Since $K_p{}^0 \times D_p{}^0$ is a characteristic subgroup of $K_p \times D_p$ this implies that $N(K_p \times D_p) \subset L \times D$. If D is a p-group and d does not divide $(2^a + 1)$ then $D_p{}^0$ is a characteristic subgroup of $K_p \times D_p$ and Lemma 5.2 once again yields that $N(K_p \times D_p) \subset L \times D$.

---

* Received October 16, 1961.

Thus if the result were false then $N(K_p \times D_p) \subset L \times D$. Hence $K_p \times D_p$ is a Sylow p-group of G. As $K_p \times D_p$ is abelian and $D_p \cap (L \times D)' = \{1\}$, Burnside's Transfer Theorem ([7], page 173) implies that G contains a normal subgroup of index p contrary to Lemma 3. 6. The proof is complete.

LEMMA 6. 2. *Let* $X = \{\lambda_i\}$ *be the set of all irreducible characters of* $N(S_2)$ *which do not contain* M *in their kernel. Then there exists a sign* $\epsilon = \pm 1$ *and a set* $\{\Lambda_i\}$ *of irreducible characters of* G *such that*

$$(\lambda_i(1)\lambda_j - \lambda_j(1)\lambda_i)^* = \epsilon(\lambda_i(1)\Lambda_j - \lambda_j(1)\Lambda_i)$$

for all λ_i, λ_j *in* X, *where* α^* *is the generalized character of* G *induced by* α.

Proof. Since QDM is a Frobenius group every character in X is induced by a character of $S_2 \times M$. Let $m_1 = [M:M']$. The result follows from [5] Theorem 2 unless M is a non-abelian r-group for some prime r and $4d^2q^2 > m_1$. In that case let M_1 be a maximal subgroup of M which is normal in $N(S_2)$. Thus $m_1 = [M':M_1]$. Let X_1 be the subset of X consisting of those characters which have M_1 in their kernel. By the argument of [5] Theorem 2 it suffices to show the existence of characters Λ_i corresponding to λ_i in X_1. Let lqd be the maximum degree of a character in X_1. Then by [5] Lemma 2. 1

$$m_1 \equiv 0 \pmod{l^2}. \tag{1}$$

The Theorem of [9] implies that it suffices to show that

$$\frac{2^a(m_1 - 1)}{qd} > 2l. \tag{2}$$

Suppose that (2) is false. Let $m_1 = 1 + 2d(2^a - 1)b$. Then

$$\frac{2^a(2^a - 1)2db}{(2^a - 1)d} \leqq 2l$$

or by (1)

$$2^ab \leqq l \leqq \sqrt{m_1}. \tag{3}$$

Hence

$$2^{2a}b^2 \leqq m_1 < 2d2^ab.$$

Thus $2^ab < 2d$. Hence $d > 2^{a-1}$. By Lemma 6. 1 this implies that $d = 2^a + 1$.

Consequently $b < \dfrac{2(2^a + 1)}{2^a} < 3$. Thus $b = 1$ or $b = 2$.

If $b=2$ then (3) yields that $2^{a+1} \leqq \sqrt{m_1}$ or

$$2^{2a+2} \leqq m_1 = 1 + 4(2^{2a} - 1) = 2^{2a+2} - 3$$

which is nonsense. If $b=1$ then $m_1 = 1 + 2(2^{2a} - 1)$, hence $m_1 \equiv 3 \pmod 4$. Thus m_1 is not a square. Now (1) implies that $l^2 \leqq \frac{m_1}{2}$. Hence

$$2^{2a} \leqq l^2 \leqq \frac{m_1}{2} = 2^{2a} - \tfrac{1}{2}.$$

which is not the case. Consequently (2) must hold and the Lemma is proved.

Lemma 6.3. *Let χ be a non-principal irreducible character of G which is real valued on M. Then $\chi(1) \geqq 2^a(m-1)$ and $\chi \neq \Lambda_i$, where Λ_i is any character defined by Lemma* 6.2.

Proof. As qdm is odd $\lambda_i \neq \overline{\lambda_i}$ for λ_i in X. This implies that $\chi \neq \Lambda_i$ for any Λ_i defined in Lemma 6.2. Thus the Frobenius reciprocity theorem implies that

(4) $$\chi|_M = b_1\rho_1 + b_2\rho_2,$$

where ρ_1 is the character of the regular representation of S_2M/M and $\rho_1 + \rho_2$ is the character of the regular representation of S_2M. Since G is simple $b_2 \neq 0$ in equation (3). Thus equation (3) yields that $\chi(1) \geqq \rho_2(1) = 2^a(m-1)$ as required.

Lemma 6.4. *The group G cannot satisfy all the assumptions stated at the beginning of this section.*

Proof. It follows from Lemma 5.13 that G is represented as a doubly transitive permutation group on the cosets of $N(S_2)$. Thus $\| 1^*_{N(S_2)} \|^2 = 2$, where $1_{N(S_2)}$ is the principal character of $N(S_2)$. A direct computation yields that if λ is a non-principal character of $N(S_2)/S_2MQ$, then $\| \lambda^* \|^2 = \| 1^*_{N(S_2)} \|^2$. Hence

(5) $$\lambda^* = \chi_1 + \chi_2,$$

where χ_1, χ_2 are distinct non-principal characters of G. Furthermore λ^* is rational valued on M. If $X_1 = \Lambda_i$ for some Λ_i defined by Lemma 6.2, then Lemma 6.3 and (5) imply that $\chi_2 = \bar{\chi}_1$. Hence

$$1 + 2^a dm = \lambda^*(1) = 2\chi_1(1),$$

which is impossible. Now Lemma 6.3 and (5) imply that

$$1+2^a m=\lambda(1)\geqq 2(m-1)2^a.$$

Thus $m\geqq 2(m-1)$ or $2\geqq m$. As m is odd this implies that $m=1$ contrary to Lemma 5.14. The proof is complete.

As was shown earlier, Lemma 6.4 is sufficient to complete the proof of Theorem 1.

CORNELL UNIVERSITY.

REFERENCES.

[9] W. Feit, "Exceptional characters," Report of the Summer Institute on Finite Groups (1960), pp. 174-180.

[10] ———, "A characterization of the simple groups $SL(2,2^a)$," *American Journal of Mathematics*, vol. 82 (1960), pp. 281-300.

ON THE CATEGORY OF SHEAVES.*

By Alex Heller and K. A. Rowe.

---

1. **Introduction.** It is well known that the category of sheaves with values in the category of abelian groups is an exact category in the sense of [3]. Now, the category of abelian groups is an exact category which satisfies the following axiom of Grothendieck: the category has arbitrary direct sums, and the direct limit of a collection of short exact sequences indexed by a directed set is a short exact sequence (see [2]). Such an exact category will be referred to as a Grothendieck category. In this paper we consider the category of sheaves with values in a Grothendieck category which possesses arbitrary direct products and a projective generator. (For the concept of a generator, see [2]). In Section 2 below, we will define a sheaf in any exact category which possesses arbitrary direct products, so that the above concept will be meaningful. We assume, that the Grothendieck category considered is abstract in the sense that the objects are not necessarily sets. The problem of the paper is to prove the following theorem:

Theorem. (a) *The category of sheaves with values in a Grothendieck category which possesses arbitrary direct products and a projective generator is exact.*

(b) *The category of sheaves as described in* (a) *above has enough injectives.*

(We split the theorem into two parts merely for organizational purposes.)

The fundamental construction in the proof is that of the sheaf generated by a presheaf. Since we dispense with sets as objects, we must necessarily adopt a new technique in this construction.

The following notations will be fixed throughout the paper. X will denote a topological space, $\mathcal{K}$ a Grothendieck category with arbitrary direct products and a projective generator K, Σ the category of presheaves over X with values in $\mathcal{K}$, and Σ' the category of sheaves over X with values in $\mathcal{K}$.

It is well known (see, for example, [1]) that Σ is an exact category. It will be seen that Σ' is a full subcategory of Σ. We begin by stating a

---

* Received December 19, 1960; revised March 1, 1962.

lemma which gives conditions under which a full subcategory of an exact category is exact.

BASIC LEMMA. *Let Σ' be a full subcategory of the exact category Σ, and let $I\colon \Sigma' \to \Sigma$ be the inclusion functor. Suppose there exists a functor $T^*\colon \Sigma \to \Sigma'$ such that IT^* is left exact and $T^*I = 1\colon \Sigma'$ (the identity functor on Σ'), and suppose there exists a natural transformation $t^*\colon (1\colon \Sigma) \to IT^*$ such that $T^*(t^*_A) = 1\colon (T^*A)$ for all objects A in Σ. Then Σ' is exact.*

Thus, to prove assertion (a) of the Theorem, we will construct a functor T^* and a natural transformation t^* with the properties described in the lemma. Of course, if F is a presheaf, T^*F will be the sheaf generated by F as described above. Our procedure will be to define a functor $T\colon \Sigma \to \Sigma$ and a natural transformation $t\colon (1\colon \Sigma) \to T$. Then the desired T^* and t^* will be a sort of "transfinite iterate" of T and t.

We begin by proving the Basic Lemma. Let

$$0 \to A' \xrightarrow{a'} A \xrightarrow{a''} A'' \to 0$$

be exact in Σ. Then we claim that

$$0 \to T^*A' \xrightarrow{T^*a'} T^*A \xrightarrow{T^*a''} T^*A'' \to 0$$

is exact in Σ'. For, let $f\colon B \to T^*A'$ with $(T^*a')f = 0$. Then $(IT^*a')(If) = 0$, and so $If = 0$ since IT^* is left exact. Hence, $f = T^*If = 0$. Now, let $f\colon B \to T^*A$ and $(T^*a'')f = 0$. Then $(IT^*a'')(If) = 0$ so that there exists a map $h\colon IB \to IT^*A'$ such that $(IT^*a')h = If$. Hence $f = T^*If = (T^*a')(T^*h)$. Next, let $f\colon T^*A'' \to B$ with $f(T^*a'') = 0$. Then $(If)(IT^*a'') = 0$, so certainly $(If)(IT^*a'')t^*_A = (If)t^*_{A''}a'' = 0$. Hence, $(If)t^*_{A''} = 0$. Then $f = T^*If = (T^*If)(T^*(t^*_{A''})) = 0$. Finally, let $f\colon T^*A \to B$ with $f(T^*a') = 0$. Then $(If)(IT^*a') = 0$, and so $(If)(IT^*a')t^*_{A'} = 0 = (If)t^*_A a'$. Hence there exists $h\colon A'' \to IB$ such that $ha'' = (If)T^*_A$. Then

$$(T^*h)(T^*a'') = (T^*If)(T^*(t^*_A)) = f.$$

This proves the exactness in question. The only non-trivial axiom to be verified is the existence of the kernel, image and cokernel of a map in Σ'. But, given a map $f\colon A \to B$ in Σ', we need only set $\ker(f) = T^*(\ker(If))$, $\operatorname{im}(f) = T^*(\operatorname{im}(If))$ and $\operatorname{coker}(f) = T^*(\operatorname{coker}(If))$. Q. E. D.

COROLLARY. *Suppose in addition that Σ has a generator R and that for each object A of Σ' we have $t^*_{IA} = 1\colon (IA)$. Then $T^*(R)$ is a generator of Σ'.*

Proof. Suppose $A \xrightarrow{f} B \to 0$ is exact in Σ' with $B \neq 0$. Then $If \neq 0$. The sequence $IA \xrightarrow{If} \operatorname{im}(If) \to 0$ is exact in Σ and $\operatorname{im}(If) \neq 0$. Hence, there exists a map $g\colon R \to IA$ such that $(If)g \neq 0$. Suppose the map $(T^{*}(If))(T^{*}g) = f(T^{*}g) = 0$. Then we have

$$0 = (If)(IT^{*}g)t^{*}_{B} = (If)t^{*}_{IA}g = (If)g \neq 0,$$

a contradiction. Q. E. D.[1]

Thus, to prove assertion (b) of the Theorem, we will find a generator R of Σ, vsrify that for a sheaf F in Σ', $t^{*}_{IF} = 1\colon (IF)$ and finally show that Σ' is a Grothendieck category. The conclusion will then follow from Theorem 1.10.1 of [2].

We now turn our attention to the definition of a sheaf in $\mathcal{K}$ and the construction of the functor T.

2. Definition of a sheaf and construction of the functor *T*. For U and V open in X, $V \subset U$, let $\rho^{U}{}_{V}\colon V \to U$ denote the natural injection. We define the category of open covers of U, denoted by $C(U)$, as follows: the objects of $C(U)$ are the open covers of U, where by "open cover of U" we mean a function $\mathcal{U}$ whose domain is U and is such that for each $x \in U$, $x \in \mathcal{U}(x) \subset U$, and $\mathcal{U}(x)$ is open in X. If $\mathcal{U}$ and $\mathcal{V}$ are two open covers of U, we say that $\mathcal{V}$ is a refinement of $\mathcal{U}$ and write $\mathcal{V} > \mathcal{U}$ or $\mathcal{U} < \mathcal{V}$ if $\mathcal{V}(x) \subset \mathcal{U}(x)$ for all $x \in U$. We set

$$\operatorname{Hom}(\mathcal{U}, \mathcal{V}) = \begin{cases} \{[\mathcal{U}, \mathcal{V}]\} & \text{if } \mathcal{V} > \mathcal{U} \\ \emptyset & \text{otherwise} \end{cases}$$

and define composition by the formula

$$[\mathcal{V}, \mathcal{W}][\mathcal{U}, \mathcal{V}] = [\mathcal{U}, \mathcal{W}] \quad \text{for } \mathcal{U} > \mathcal{U} > \mathcal{W}.$$

Given a presheaf F in Σ, and an open cover $\mathcal{U}$ of the open set U, we consider the sequence

$$(1) \quad 0 \to F(U) \xrightarrow{\theta[F, U, \mathcal{U}]} \prod_{x \in U} F(\mathcal{U}(x)) \xrightarrow{\psi[F, U, \mathcal{U}]} \prod_{x, y \in U} F(\mathcal{U}(x) \cap \mathcal{U}(y)).$$

The maps in this sequence are defined as follows:

---

[1] We note in passing that under the hypotheses of the Corollary, the maps $\operatorname{Hom}(A, IB) \to \operatorname{Hom}(T^{*}A, B)$ given by $f \to T^{*}f$ and $\operatorname{Hom}(T^{*}A, B) \to \operatorname{Hom}(A, IB)$ given by $g \to (Ig)t^{*}_{A}$ are isomorphisms. Hence T^{*} and t^{*} solve the universal factorization problem for maps of presheaves into sheaves.

$$(p_x[F, U, \mathcal{U}])\,(\theta[F, U, \mathcal{U}]) = F(\rho^{U}_{\mathcal{U}(x)})$$

$$\begin{aligned}(p_{x,y}[F, U, \mathcal{U}])\,(\psi[F, U, \mathcal{U}]) = {} & F(\rho^{\mathcal{U}(x)}_{\mathcal{U}(x)\cap\mathcal{U}(y)})\,(p_x[F, U, \mathcal{U}]) \\ & - F(\rho^{\mathcal{U}(y)}_{\mathcal{U}(x)\cap\mathcal{U}(y)})\,(p_y[F, U, \mathcal{U}])\end{aligned}$$

where $x, y \in U$, and where the p's denote the projections of the direct products in (1). It is easily seen that $\psi[F, U, \mathcal{U}]\theta[F, U, \mathcal{U}] = 0$, but that the sequence is not necessarily exact. We define a sheaf as follows:

Definition. A presheaf F is a sheaf if and only if the sequence (1) is exact for all open sets U and open covers $\mathcal{U}$.

With this definition it is easily seen that the collection of all sheaves and natural transformations between sheaves constitute a full subcategory of the category Σ.

We now begin the construction of T. We define a direct system of objects and maps in $\mathcal{K}$, indexed by $C(U)$, for each choice of a presheaf F and an open set U. This is equivalent to defining a (covariant) functor $F_U\colon C(U) \to \mathcal{K}$. If $\mathcal{U}$ is an open cover of U, we set

$$F_U(\mathcal{U}) = \ker(\psi[F, U, \mathcal{U}]).$$

To define the map $F_U([\mathcal{U}, \mathcal{V}])\colon F_U(\mathcal{U}) \to F_U(\mathcal{V})$ for $\mathcal{V} > \mathcal{U}$, we consider the diagram:

$$(2)\qquad \begin{tikzcd} 0 \arrow[r] & F_U(\mathcal{U}) \arrow[r, "{\lambda[F, U, \mathcal{U}]}"] & \prod F(\mathcal{U}(x)) \arrow[r, "{\psi[F, U, \mathcal{U}]}"] \arrow[d, "{\sigma[F, U, \mathcal{U}, \mathcal{V}]}"] & \prod F(\mathcal{U}(x) \cap \mathcal{U}(y)) \arrow[d, "{\tau[F, U, \mathcal{U}, \mathcal{V}]}"] \\ 0 \arrow[r] & F_U(\mathcal{V}) \arrow[r, "{\lambda[F, U, \mathcal{V}]}"] & \prod F(\mathcal{V}(x)) \arrow[r, "{\psi[F, U, \mathcal{V}]}"] & \prod F(\mathcal{V}(x) \cap \mathcal{V}(y)) \end{tikzcd}$$

The rows of this diagram are exact by definition. The vertical maps are defined by the formulas

$$p_x[F, U, \mathcal{V}]\sigma[F, U, \mathcal{U}, \mathcal{V}] = F(\rho^{\mathcal{U}(x)}_{\mathcal{V}(x)})\,p_x[F, U, \mathcal{U}]$$

$$p_{x,y}[F, U, \mathcal{V}]\tau[F, U, \mathcal{U}, \mathcal{V}] = F(\rho^{\mathcal{U}(x)\cap\mathcal{U}(y)}_{\mathcal{V}(x)\cap\mathcal{V}(y)})\,p_{x,y}[F, U, \mathcal{U}]$$

where $x, y \in U$. It is easily seen that diagram (2) commutes, and so defines the required map. Further, if $\mathcal{W} > \mathcal{V} > \mathcal{U}$, we see, by considering diagram (2) with a third row corresponding to the cover $\mathcal{W}$, that

$$F_U([\mathcal{V}, \mathcal{W}])F_U([\mathcal{U}, \mathcal{V}]) = F_U([\mathcal{U}, \mathcal{W}]).$$

Thus, F_U is a functor.

Since $\mathcal{K}$ has arbitrary direct sums, the functor F_U has a direct limit. We set

$$TF(U) = \underset{\mathcal{U}}{\underrightarrow{\lim}} \; F_U(\mathcal{U}). \tag{3}$$

Let $h[F, U, \mathcal{U}] : F_U(\mathcal{U}) \to TF(U)$ denote the injections into the limit.

Next, we show that TF is a presheaf, where for U open in X, $TF(U)$ is defined by (3). We must define $TF(\rho^U{}_V) : TF(U) \to TF(V)$ for $V \subset U$, and show that the maps so defined satisfy the requirements. To do this, if $\mathcal{U}$ is an open cover of U, we define the open cover of V, $\mathcal{U} \mid V$, by the formula $(\mathcal{U} \mid V)(x) = \mathcal{U}(x) \cap V$ for $x \in V$. We now consider the diagram:

$$\begin{tikzcd}
0 \arrow[r] & F_U(\mathcal{U}) \arrow[r, "{\psi[F,U,\mathcal{U}]}"] & \prod F(\mathcal{U}(x)) \arrow[r, "{\lambda[F,U,\mathcal{U}]}"] \arrow[d, "{\xi[F,U,V,\mathcal{U}]}"] & \prod F(\mathcal{U}(x) \cap \mathcal{U}(y)) \arrow[d, "{\eta[F,U,V,\mathcal{U}]}"] \\
& F_V(\mathcal{U} \mid V) \arrow[r, "{\lambda[F,V,\mathcal{U} \mid V]}"] & \prod F(\mathcal{U}(x) \cap V) \arrow[r, "{\psi[F,V,\mathcal{U} \mid V]}"] & \prod F(\mathcal{U}(x) \cap \mathcal{U}(y) \cap V)
\end{tikzcd}$$

where the projections of the vertical maps are defined in terms of the appropriate restriction maps. Again, the diagram commutes, and so defines a map $\zeta[F, U, V, \mathcal{U}] : F_U(\mathcal{U}) \to F_V(\mathcal{U} \mid V)$ which commutes with the rest of the diagram. Now we consider the map $g[F, U, V, \mathcal{U}] = h[F, V, \mathcal{U} \mid V]\, \zeta[F, U, V, \mathcal{U}]$. It is easily seen that if $\mathcal{V} > \mathcal{U}$, we have $g[F, U, V, \mathcal{U}]\; F_U([\mathcal{U}, \mathcal{V}]) = g[F, U, V, \mathcal{U}]$. Hence, by the universal property of direct limits, there exists a map $TF(\rho^U{}_V) : TF(U) \to TF(V)$ such that $TF(\rho^U{}_V) h[F, U, \mathcal{U}] = g[F, U, V, \mathcal{U}]$. We obviously have $TF(\rho^U{}_U) = 1 : (TF(U))$, and it is a straightforward computation to verify that $TF(\rho^V{}_W) TF(\rho^U{}_V) = TF(\rho^U{}_W)$ if $W \subset V \subset U$. Hence, TF is a presheaf.

We are now in a position to define T. Of course, on objects, i.e., on presheaves, T is defined by the above constructions. It remains to define T on maps. Thus, let $f : F \to G$ be a map of presheaves. For U open in X and $\mathcal{U}$ an open cover of U, consider the diagram

$$\begin{tikzcd}
0 \arrow[r] & F_U(\mathcal{U}) \arrow[r, "{\lambda[F,U,\mathcal{U}]}"] & \prod F(\mathcal{U}(x)) \arrow[r, "{\psi[F,U,\mathcal{U}]}"] \arrow[d, "{r[f,U,\mathcal{U}]}"] & \prod F(\mathcal{U}(x) \cap \mathcal{U}(y)) \arrow[d, "{s[f,U,\mathcal{U}]}"] \\
0 \arrow[r] & G_U(\mathcal{U}) \arrow[r, "{\lambda[G,U,\mathcal{U}]}"] & \prod G(\mathcal{U}(x)) \arrow[r, "{\psi[G,U,\mathcal{U}]}"] & \prod G(\mathcal{U}(x)) \cap \mathcal{U}(y))
\end{tikzcd}$$

where the projections of the vertical maps are defined in terms of the appropriate maps given by f. This diagram commutes and so defines a map $q[f, U, \mathfrak{U}] : F_U(\mathfrak{U}) \to G_U(\mathfrak{U})$ which commutes with the rest of the diagram. Then the map $k[f, U, \mathfrak{U}] = h[G, U, \mathfrak{U}]\, q[f, U, \mathfrak{U}]$ satisfies the condition that $k[f, U, \mathfrak{V}] F_U([\mathfrak{U}, \mathfrak{V}]) = k[f, U, \mathfrak{U}]$ for $\mathfrak{V} > \mathfrak{U}$, and so there exists a map $(Tf)_U : TF(U) \to TG(U)$ such that $(Tf)_U h[F, U, \mathfrak{U}] = k[f, U, \mathfrak{U}]$. Finally, we observe that $(Tf)_V TF(\rho^U{}_V) = TG(\rho^U{}_V)(Tf)_U$ if $V \subset U$. Hence, Tf: $TF \to TG$ defined by the above constructions is a natural transformation, i. e., a map of presheaves.

To complete the construction of T, it remains to show that if $f: F \to F'$ and $f' : F' \to F''$, then $T(f' \circ f) = (Tf')(Tf)$. This is a matter of routine computation.

3. **The natural transformation t:$(1 : \Sigma) \to T$.** Let F denote a presheaf, and $\mathfrak{U}$ an open cover of the open set U. There exists a map $\mu[F, U, \mathfrak{U}] : F(U) \to F_U(\mathfrak{U})$ such that $\lambda[F, U, \mathfrak{U}]\, \mu[F, U, \mathfrak{U}] = \theta[F, U, \mathfrak{U}]$. Consider the map $t[F, U, \mathfrak{U}] = h[F, U, \mathfrak{U}]\, \mu[F, U, \mathfrak{U}]$. We show that this map is independent of $\mathfrak{U}$. For, let $\mathfrak{V}$ be any other cover of U, and consider the cover $\mathfrak{W}$ given by $\mathfrak{W}(x) = \mathfrak{U}(x) \cap \mathfrak{V}(x)$. Then $\mathfrak{W} > \mathfrak{U}$ and $\mathfrak{W} > \mathfrak{V}$. It is easily seen that $F_U([\mathfrak{X}, \mathfrak{W}])\mu[F, U, \mathfrak{X}] = \mu[F, U, \mathfrak{W}]$, where $\mathfrak{X}$ is either $\mathfrak{U}$ or $\mathfrak{V}$. It now follows immediately that $t[F, U, \mathfrak{U}] = t[F, U, \mathfrak{V}]$. Hence, we can define a map $t_{F,U}$: $F(U) \to TF(U)$. This in turn defines a natural transformation t_F: $F \to TF$. Indeed, let $V \subset U$, and let $\mathfrak{U}$ be an open cover of U. Then the desired property follows immediately from the relation

$$p_x[F, V, \mathfrak{U} \mid V]\lambda[F, V, \mathfrak{U} \mid V]\,\{\zeta[F, U, V, \mathfrak{U}]\mu[F, U, \mathfrak{U}]$$
$$-\mu[F, V, \mathfrak{U} \mid V]F(\rho_V{}^U)\} = 0$$

for any $x \in V$. Finally, this defines a natural transformation t: $(1 : \Sigma) \to T$. For, if f: $F \to G$ is a map of presheaves and $\mathfrak{U}$ is an open cover of the open set U, then the formula $(Tf)_U t_{F,U} = t_{G,U} f_U$ follows from the relation

$$p_x[G, U, \mathfrak{U}]\,\lambda[G, U, \mathfrak{U}]\,\{q[f, U, \mathfrak{U}]\,\mu[F, U, \mathfrak{U}] - \mu[G, U, \mathfrak{U}]f_U\} = 0$$

for any $x \in U$.

4. **The functor T^* and the natural transformation t^*.** We now proceed to construct T^*: $\Sigma \to \Sigma$. However, it will be shown in Section 5 that if F is a presheaf, then T^*F is a sheaf. Since T^* will be constructed to be a functor, we may then regard it as a functor $\Sigma \to \Sigma'$. Simultaneously we will

construct t^*, and in Section 6, it will be shown that T^* and t^* satisfy the hypothesis of the Basic Lemma, thus proving the first assertion of the theorem.

First, we note the following: if $L: \Sigma \to \Sigma$ is any functor, then the map $t(L): L \to TL$ defined by $(t(L))_F = t_{L(F)}$ is a natural transformation.

Let Γ denote the cardinal number of the set of all subgadgets of the generator K of the category $\mathcal{K}$. Let ϵ denote a non-limit ordinal such that $\epsilon > \omega$ and $\aleph_\epsilon > \Gamma$. Let δ denote the first ordinal of cardinality $\aleph_\epsilon$. We recall that δ is then a "regular ordinal" (see, for example, [4], pg. 225). In essence, T^* may be thought of as the "δ-th iterate" of T. We now make this concept precise.

For each $\beta \leqq \delta$ let C_β denote the category of all ordinals less than β. We will denote the set $\mathrm{Hom}(\alpha, \alpha')$ for $\alpha \leqq \alpha'$ in the category C_β by (α, α'). We proceed to define a functor $P: C_\delta \to \mathrm{Hom}(\Sigma, \Sigma)$ — category of all functors from Σ to Σ, with natural transformations as maps. P will be defined by transfinite induction. First, we set $P(0) = 1: \Sigma$, $P(1) = T$, and $P(0,1) = t$. Now suppose that for some β, $0 < \beta < \delta$, P is defined on C_β. We then define P on $C_{\beta+1}$. First, suppose $\beta = \beta' + 1$. Then we write $P(\beta) = T \cdot P(\beta')$, $P(\alpha, \beta) = t(P(\beta')) \circ P(\alpha, \beta')$ for $\alpha < \beta$, and $P(\beta, \beta) = 1: (T \cdot P(\beta'))$. On the other hand, if $\beta = \lim\limits_{\alpha<\beta} \alpha$, we set $P(\beta) = \varinjlim\limits_{\alpha<\beta} P(\alpha)$, and let $P(\alpha, \beta)$ be the injections into the direct limit for $\alpha < \beta$. In this case we let $P(\beta, \beta) = 1: (\varinjlim\limits_{\alpha<\beta} P(\alpha))$. We note that the direct limits exist since $\mathcal{K}$ has arbitrary direct sums, hence the same is true of Σ, and so also of $\mathrm{Hom}(\Sigma, \Sigma)$.

It now follows by transfinite induction that P is defined on the category C_δ. We define

$$T^* = \varinjlim_{\alpha<\delta} P(\alpha).$$

For $\alpha < \delta$ let t^α denote the injection $P(\alpha) \to T^*$. We define

$$t^* = t^0.$$

We also write $T^\alpha = P(\alpha)$, and $t^\alpha{}_\beta = P(\alpha, \beta)$ for $\alpha \leqq \beta < \delta$.

5. The sheaf generated by a presheaf. Let F denote a presheaf in Σ. In this section we prove the following lemma:

PRINCIPAL LEMMA. *T^*F is a sheaf.*

We prove this by showing that if $\mathcal{U}$ is an open cover of the open set U, then the sequence (1), with F replaced by T^*F, is exact.

Throughout this section we will keep F, U, and $\mathcal{U}$ fixed. We first simplify the notation. For $\alpha < \delta$, set $A'_\alpha = T^\alpha F(U)$, $A_\alpha = \prod T^\alpha F(\mathfrak{U}(x))$, $A''_\alpha = \prod T^\alpha F(\mathfrak{U}(x) \cap \mathfrak{U}(y))$, $\theta_\alpha = \theta[T^\alpha F, U, \mathfrak{U}]$, and $\psi_\alpha = \psi[T^\alpha F, U, \mathfrak{U}]$. For $\alpha \leqq \beta$, set $a'_{\alpha,\beta} = (t^\alpha{}_\beta)_{F,U}$, and define maps $a_{\alpha,\beta}\colon A_\alpha \to A_\beta$ and $a''_{\alpha,\beta}\colon A''_\alpha \to A''_\beta$ in terms of the appropriate t's. Next, set $A' = T^*F(U)$, $A = \prod T^*F(\mathfrak{U}(x))$, $A'' = \prod T^*F(\mathfrak{U}(x) \cap \mathfrak{U}(y))$, $\theta = \theta[T^*F, U, \mathfrak{U}]$, and $\psi = \psi[T^*F, U, \mathfrak{U}]$. Finally, for $\alpha < \delta$, set $a'_\alpha = (t^\alpha)_{F,U}$, and define maps $a_\alpha\colon A_\alpha \to A$ and $a''_\alpha\colon A''_\alpha \to A''$ in terms of the appropriate t's. It is easily seen that $a_{\beta,\gamma}a_{\alpha,\beta} = a_{\alpha,\gamma}$ and $a''_{\beta,\gamma}a''_{\alpha,\beta} = a''_{\alpha,\gamma}$ if $\alpha \leqq \beta \leqq \gamma < \delta$, and that A and A'' are direct limits of the A_α and A''_α respectively with a_α and a''_α as injections.

We precede the proof of the Principal Lemma by a definition and some lemmas pertaining to the category $\mathcal{K}$.

Definition. Consider the commutative diagram (in $\mathcal{K}$):

$$\begin{tikzcd}
0 \arrow[r] & B' \arrow[r, "b'"] \arrow[d, "h'"] & B \arrow[r, "b''"] \arrow[d, "h"] & B'' \arrow[d, "h''"] \\
0 \arrow[r] & D' \arrow[r, "d'"] & D \arrow[r, "d''"] & D''
\end{tikzcd}$$

Suppose that $b''b' = 0$ and $d''d' = 0$. We say that the second row of this diagram is relatively exact with respect to the first if and only if (1) given $f\colon M \to B'$ with $b'f = 0$, then $h'f = 0$; (2) given $f\colon M \to B$ with $b''f = 0$, then there exists a map $g\colon M \to D'$ such that $d'g = hf$.

LEMMA 1. *Suppose $\{B_\alpha, b^\alpha{}_\beta\}$ is a direct system of objects and maps in $\mathcal{K}$ indexed by the set $\lambda \leqq \alpha < \delta$. Let $B = \varinjlim_{\lambda\leqq\alpha<\delta} B_\alpha$, and let $b^\alpha\colon B_\alpha \to B$ denote the injections. Suppose also that for $\lambda \leqq \alpha < \delta$ we have maps f_α: $K \to B_\alpha$ such that $f_\beta = b^\alpha{}_\beta f_\alpha$ for $\lambda \leqq \alpha \leqq \beta < \delta$. Finally, suppose that $f = b^\lambda f_\lambda = 0$. Then there exists a $\mu \geqq \lambda$ such that $f_\mu = 0$.*

Proof. For each $\alpha \geqq \lambda$ we factor f_α obtaining exact sequences

$$0 \to K_\alpha \xrightarrow{u_\alpha} K \xrightarrow{v_\alpha} M_\alpha \to 0 \text{ and } 0 \to M_\alpha \xrightarrow{w_\alpha} B_\alpha,$$

where $w_\alpha v_\alpha = f_\alpha$. For $\lambda \leqq \alpha \leqq \beta < \delta$, $w_\beta v_\beta u_\alpha = f_\beta u_\alpha = b^\alpha{}_\beta f_\alpha u_\alpha = 0$, so that $v_\beta u_\alpha = 0$. Hence, there exists maps $k^\alpha{}_\beta\colon K_\alpha \to K_\beta$ and $m^\alpha{}_\beta\colon M_\alpha \to M_\beta$ such that $u_\beta k^\alpha{}_\beta = u_\alpha$, $m^\alpha{}_\beta v_\alpha = v_\beta$, $w_\beta m^\alpha{}_\beta = b^\alpha{}_\beta w_\alpha$ and the systems $\{K_\alpha, k^\alpha{}_\beta\}$ and $\{M_\alpha, m^\alpha{}_\beta\}$ are direct. Taking the direct limit of the above sequences we obtain

$0 \to \varinjlim K_\alpha \xrightarrow{u} K \xrightarrow{v} \varinjlim M_\alpha \to 0$ and $0 \to \varinjlim M_\alpha \to B$, which are exact since $\mathcal{K}$ is a Grothendieck category. If $m^\alpha: M_\alpha \to \varinjlim M_\alpha$ are the injections then for $\alpha \geqq \lambda$ we have $wm^\alpha v_\alpha = b^\alpha w_\alpha v_\alpha = b^\alpha f_\alpha = 0$. Hence, $w = 0$, and we conclude $K = \varinjlim K_\alpha$. Because of the cardinality of δ and the fact that it is a regular ordinal, there exists a $\mu \geqq \lambda$ such that if $\alpha \geqq \mu$ then $K^\alpha = K_\mu$. Thus $\varinjlim K_\alpha = K_\mu$. Hence $M_\mu = 0$, i.e., $f_\mu = 0$. Q.E.D.

LEMMA 2. *Let* λ, B, $b^\alpha{}_\beta$, B_α *and* b^α *be as in the previous lemma. Suppose* $f: K \to B$. *Then there exists* $\mu \geqq \lambda$ *and a map* $g_\mu: K \to B_\mu$ *such that* $b^\mu g_\mu = f$.

Proof. For each $\alpha \geqq \lambda$ consider the exact sequence $B_\alpha \xrightarrow{b^\alpha} B \xrightarrow{d^\alpha} C_\alpha \to 0$ obtained by factoring b^α. Set $f_\alpha = d^\alpha f$. Now if $\alpha \leqq \beta$ then $d^\beta b^\alpha = d^\beta b^\beta b^\alpha{}_\beta = 0$, so that there exist maps $c^\alpha{}_\beta: C_\alpha \to C_\beta$ such that $c^\alpha{}_\beta d^\alpha = d^\beta$ and the system $\{C_\alpha, c^\alpha{}_\beta\}$ is direct. But $\varinjlim_{\lambda \leqq \alpha < \delta} C_\alpha = 0$. Hence $c^\alpha f_\alpha = 0$ for all $\alpha \geqq \lambda$, where the c^α are the injections. By Lemma 1, there exists $\mu \geqq \lambda$, such that $f_\mu = d^\mu f = 0$. Since K is projective, there exists $g_\mu: K \to B_\mu$ such that $b^\mu g_\mu = f$. Q.E.D.

LEMMA 3. *Suppose that for each* $\alpha < \delta$ *we are given a sequence*

$$0 \to B'_\alpha \xrightarrow{b'_\alpha} B_\alpha \xrightarrow{b''_\alpha} B''_\alpha$$

in K, *and for* $\alpha \leqq \beta < \delta$ *maps* $d'_{\alpha\beta}: B'_\alpha \to B'_\beta$, $d_{\alpha\beta}: B_\alpha \to B_\beta$, $d''_{\alpha\beta}: B''_\alpha \to B''_\beta$ *such that the resulting systems are direct. Let* $B' = \varinjlim_{\alpha<\beta} B'_\alpha$, *etc., and let* d'_α, d_α, d''_α *denote the injections, and* b', b'' *be the limits of the maps* b'_α, b''_α. *Suppose that these maps give rise to commutative diagrams. Finally suppose that if* $\beta > \alpha$ *the sequence with index* β *is relatively exact with respect to the sequence with index* α. *Then the sequence* $0 \to B' \xrightarrow{b'} B \xrightarrow{b''} B''$ *is exact.*

Proof. We observe that it is sufficient to show that the sequence $0 \to \operatorname{Hom}(K, B') \to \operatorname{Hom}(K, B) \to \operatorname{Hom}(K, B'')$ is exact. First, it is trivial that $b''b' = 0$. Suppose then that $f: K \to B'$ with $b'f = 0$. By Lemma 2, there exists a $\lambda \geqq 0$ and a map $g_\lambda: K \to B'_\lambda$ such that $d'_\lambda g_\lambda = f$. Then $0 = b'f = b'd'_\lambda g_\lambda = d_\lambda b'_\lambda g_\lambda$. For $\alpha \geqq \lambda$ set $f_\alpha = b'_\alpha d'_{\lambda\alpha} g_\lambda$. Then $d_\alpha f_\alpha = 0$ for $\alpha \geqq \lambda$. By Lemma 1, there exists $\mu \geqq \lambda$ such that $f_\mu = 0$. Then if $\alpha > \mu$, $d'_{\mu\alpha} d'_{\lambda\mu} g_\lambda = 0$. Hence, $f = d'_\lambda g_\lambda = d'_\alpha d'_{\mu\alpha} d'_{\lambda\mu} g_\lambda = 0$.

Now suppose that $f: K \to B$ with $b''f = 0$. By Lemma 2, there exists a $\lambda \geqq 0$ and a map $g_\lambda: K \to B_\lambda$ such that $d_\lambda g_\lambda = f$. For $\alpha \geqq \lambda$ set $g_\alpha = d_{\lambda\alpha} g_\lambda$.

Then $d''_\alpha b''_\alpha g_\alpha = b''f = 0$. Hence, by Lemma 1, there exists $\mu \geqq \lambda$ such that $b''_\mu g_\mu = 0$. Hence, if $\alpha > \mu$, there exists $f'_\alpha \colon K \to B'_\alpha$ such that $b'_\alpha f'_\alpha = d_{\mu\alpha} g_\mu$. Set $g = d'_\alpha f'_\alpha$. Then $b'g = f$. Q. E. D.

In view of these lemmas, the Principal Lemma will be proved if for $\beta > \alpha$ the sequence $0 \to A'_\beta \to A_\beta \to A''_\beta$ is relatively exact with respect to the sequence $0 \to A'_\alpha \to A_\alpha \to A''_\alpha$. We prove this as a separate lemma.

LEMMA 4. *If $\beta > \alpha$ the sequence (1) with $T^\beta F$ in place of F is relatively exact with respect to the sequence with $T^\alpha F$ in place of F.*

Proof. We show first that the diagram

$$\begin{tikzcd}
0 \arrow[r] & (T^\alpha F)_U(\mathcal{U}) \arrow[rr, "{\lambda[T^\alpha F, U, \mathcal{U}]}"] \arrow[d, "{h[T^\alpha F, U, \mathcal{U}]}"] & & A_\alpha \arrow[r, "\psi_\alpha"] \arrow[d, "a_{\alpha,\alpha+1}"] & A''_\alpha \arrow[d, "a''_{\alpha,\alpha+1}"] \\
0 \arrow[r] & A'_{\alpha+1} \arrow[rr, "\theta_{\alpha+1}"] & & A_{\alpha+1} \arrow[r, "\psi_{\alpha+1}"] & A''_{\alpha+1}
\end{tikzcd}$$

in which the first row is exact, commutes for every α. The right-hand square is obviously commutative by definition of the symbols. To prove the commutativity of the left-hand square, we have merely to observe that

$$\begin{aligned}(t^\alpha{}_{\alpha+1})_{F,\mathcal{U}(x)} &= t_{T^\alpha F,\mathcal{U}(x)} \\ &= h[T^\alpha F, \mathcal{U}(x), \mathcal{U} \mid \mathcal{U}(x)]\mu[T^\alpha F, \mathcal{U}(x), \mathcal{U} \mid \mathcal{U}(x)]\end{aligned}$$

for $x \in U$, and then that for

$$\begin{aligned} y \in \mathcal{U}(x), \{ & T^\alpha F(\rho^{\mathcal{U}(y)}_{\mathcal{U}(x) \cap \mathcal{U}(y)}) p_y[T^\alpha F, U, \mathcal{U}] \\ & - T^\alpha F(\rho^{\mathcal{U}(x)}_{\mathcal{U}(x) \cap \mathcal{U}(y)}) p_x[T^\alpha F, U, \mathcal{U}]\} \lambda[T^\alpha F, U, \mathcal{U}] = 0 \end{aligned}$$

by definition.

To prove the lemma, suppose first that $f \colon M \to A'_\alpha$ with $\theta_\alpha f = 0$. Then $\lambda[T^\alpha F, U, \mathcal{U}]\mu[T^\alpha F, U, \mathcal{U}]f = 0$, so that $\mu[T^\alpha F, U, \mathcal{U}]f = 0$. Thus $a'_{\alpha,\alpha+1}f = h[T^\alpha F, U, \mathcal{U}]\mu[T^\alpha F, U, \mathcal{U}]f = 0$, so finally $a'_{\alpha\beta}f = a'_{\alpha+1,\beta}a'_{\alpha,\alpha+1}f = 0$. Now suppose $f \colon M \to A_\alpha$ with $\psi_\alpha f = 0$. Then there exists a map $g^* \colon M \to (T^\alpha F)_U(\mathcal{U})$ such that $\lambda[T^\alpha F, U, \mathcal{U}]g^* = f$. Set $g = a'_{\alpha+1,\beta}h[T^\alpha F, U, \mathcal{U}]g^*$. Then we have $\theta_\beta g = a_{\alpha,\beta}f$. Q. E. D.

This completes the proof of the Principal Lemma. Naturally, T^*F is the sheaf generated by the presheaf F.

6. Proof of assertion (a) of the theorem. To prove assertion (a) we have only to verify that T^* and t^* satisfy the conditions of the Basic Lemma of the introduction. As noted above, we regard T^* as having values in Σ', not Σ.

First, we observe that if F is already a sheaf, $TF = F$ and hence $T^*F = F$, showing that $T^*I = 1 : \Sigma'$. For in this case the sequence (1) is exact, and so $F(U) = F_U(\mathcal{U})$. Thus $TF(U) = \varinjlim F_U(\mathcal{U}) = F(U)$. The Grothendieck axiom in $\mathcal{K}$ makes IT^* left exact. Finally the formula $T^*(t^*_F) = 1 : (T^*F)$ follows immediately from the fact that T^*F is a sheaf. Thus, the conditions are satisfied, and we conclude that the category Σ' is exact.

7. Proof of assertion (b) of the theorem. (1) Definition of a generator R of Σ.

We define R by setting $R(U) = \sum_{V \supset U} K$. Let $(i_U)_V : K \to R(U)$ denote the injections of K into the direct sum. Then of course, if $U \supset U'$ we define $R(\rho^U{}_{U'})$ by $R(\rho^U{}_{U'})(i_U)_V = (i_{U'})_V$ for any $V \supset U$. To see that R is a generator let $A \xrightarrow{\Phi} B \to 0$ be an exact sequence of presheaves with $B \neq 0$. For each open set U such that $B(U) \neq 0$ choose a map $f_U : K \to A(U)$ such that $\Phi_U f_U \neq 0$. For each open set U such that $B(U) = 0$ define $f_U : K \to A(U)$ to be zero. To define a natural transformation $\psi : R \to A$ we proceed as follows: let U be open, and define ψ_U by $\psi_U(i_U)_V = A(\rho_U{}^V) f_V$ for any $V \supset U$. That ψ is a natural transformation is immediate. We claim $\Phi\psi \neq 0$. Suppose $\Phi\psi = 0$. Then $\Phi_U\psi_U = 0$ for all open sets U. Now choose an open set U such that $B(U) \neq 0$. Then $\Phi_U\psi_U(i_U)_V = 0$ for all $V \supset U$. In particular, we have

$$0 = \Phi_U\psi_U(i_U)_U = \Phi_U A(\rho^U{}_U) f_U = \Phi_U f_U \neq 0,$$

a contradiction.

(2) For F a sheaf, the relation $t^*_{IF} = 1 : (IF)$ is immediate, since, in this case $\mu[IF, U, \mathcal{U}] = h[IF, U, \mathcal{U}] = 1 : (IF(U))$ for any open U, and any cover $\mathcal{U}$ of U, showing that $(t_{IF})_U = 1 : (IF(U))$ and hence that $P(\alpha, \beta) = 1 : (IF)$ for $0 \leqq \alpha \leqq \beta < \delta$.

We conclude, according to the corollary to the Basic Lemma, that T^*R is a generator of Σ'.

(3) It remains to verify that Σ' is a Grothendieck category. It is immediate that Σ' has arbitrary direct sums, if we observe that $\Sigma_\alpha F_\alpha = T^*(\Sigma_\alpha(IF_\alpha))$, where $\Sigma_\alpha(IF_\alpha)$ is the presheaf defined by $U \to \Sigma_\alpha((IF_\alpha)(U))$.

Let $\{F^\alpha, f^\alpha{}_\beta\}$ denote a direct system of sheaves indexed by the directed set $\mathfrak{A}$. We define, as usual, the direct limit of this system by the formula $\varinjlim_\alpha F_\alpha = T^*(\varinjlim_\alpha (IF_\alpha))$, where $(\varinjlim_\alpha (IF_\alpha))(U) = \varinjlim_\alpha ((IF_\alpha)(U))$.

Suppose also that $\{G_\alpha, g^\alpha{}_\beta\}$ and $\{H_\alpha, h^\alpha{}_\beta\}$ are direct systems of sheaves and that we have maps c^α, d^α so that the sequences $0 \to F_\alpha \xrightarrow{c^\alpha} G_\alpha \xrightarrow{d^\alpha} H_\alpha \to 0$ are exact and the c^α, d^α commute with the maps of the direct systems. We show that the limit of these sequences is exact.

First, since every direct limit is right exact, the sequence

$$\varinjlim F_\alpha \to \varinjlim G_\alpha \to \varinjlim H_\alpha \to 0$$

is exact. Next, we note that I is left exact. (This depends on the property $t^*{}_{IF} = 1 : (IF)$ for F a sheaf.) Hence the sequences $0 \to IF_\alpha \to IG_\alpha \to IH_\alpha$ are exact, and since Σ is a Grothendieck category, the sequence

$$0 \to \varinjlim_\alpha (IF_\alpha) \to \varinjlim_\alpha (IG_\alpha) \to \varinjlim_\alpha (IH_\alpha)$$

is exact. Finally, since, by the proof of the Basic Lemma, T^* is exact, the sequence

$$0 \to \varinjlim_\alpha F_\alpha \to \varinjlim_\alpha G_\alpha \to \varinjlim_\alpha H_\alpha$$

is exact.

Hence, Σ' is a Grothendieck category, and so assertion (b) of the Theorem follows from Th. 1.10.1 of [2].

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] R. Godement, R., *Topologie Algebrique et Theorie des Faisceaux*, Paris, Hermann et Cie., 1958.

[2] A. Grothendieck, "Sur quelques points d'algebra homologique," *Tohoku Mathematical Journal*, 1957.

[3] A. Heller, "Homological algebra in Abelian categories," *Annals of Mathematics*, 1958.

[4] W. Sierpinski, *Leçons sur les Nombres Transfinies*, Paris, Gauthier-Villars et Cie., 1928.

INTERSECTION THEORY ON QUOTIENTS OF ALGEBRAIC VARIETIES.*

By R. E. Briney.[1]

Introduction. Given two cycles Z, Z' carried by an ambient algebraic variety X, we can assign to each component C of their intersection which is simple on X a well-defined multiplicity $(C; Z \cdot Z')_X$. However, no general definition has yet been given for multiplicities of intersection at components which are *singular* on the ambient variety. Certain special cases have been studied: for example, if X is projectively embedded, Samuel ([11], ch. II, 6, no. 5) has defined intersection multiplicities for those cycles which locally are complete intersections of X with cycles in the projective space. This applies in particular to the intersection of two generators of a quadratic cone in 3-space.

In the present paper, we study another particular case of this problem. Let a finite group $\mathfrak{g}$ act as a group of biregular transformations on a nonsingular variety X, and assume that the quotient space $Y = X/\mathfrak{g}$ is a variety. If elements of $\mathfrak{g}$ other than the identity have fixed points on X, the variety Y will in general have singularities. We shall define an intersection multiplicity $(C; Z \cdot Z')_Y$ for any component C of the proper dimension, $\dim Z + \dim Z' - \dim Y$; when C is simple on Y, our definition will coincide with the usual one. We shall in fact establish a global intersection theory, not only on quotients of nonsingular varieties but on varieties of a possibly more general type—those which result from nonsingular varieties after a finite number of repetitions of the process of forming quotients. The precise result is stated as Theorem 3.7.

In defining intersection multiplicities on quotient varieties, we shall be forced to use cycles with rational coefficients and to allow rational multiplicities. The example of the two generators on a quadratic cone ([11], loc. cit.)

---

* Received May 24, 1961; revised April 9, 1962.

[1] The contents of this paper are contained, in slightly different form, in a doctoral disertation presented to the Massachusetts Institute of Technology, May, 1961. The research was carried out under the direction of Professor Arthur Mattuck, to whom I wish to express my sincere gratitude. His encouragement and advice were invaluable during both the writing of the thesis and preparation of this paper. I would also like to thank R. E. Mosher for several stimulating conversations.

Part of the work on this paper was supported by the United States Air Force, through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18 (603)-90.

has already shown that one cannot expect intersection multiplicities at singular components to be integers. At the end of Section 3 we give some examples of rational intersection multiplicities on quotient varieties.

In order to carry out the proof of the above result, it is efficient to formulate axiomatically those properties of varieties which will guarantee that a global intersection theory can be established. We give these axioms in Section 1. In Section 3 we show (Theorem 3.1) that if these axioms are satisfied by the varieties in a given category, they are also satisfied by quotients of the varieties in this category. In Section 2 we recall the definition of a quotient of an alegbraic variety, and prove some preliminary results which are needed in Section 3.

The proof that the axioms listed in Section 1 are sufficient for the establishment of a global intersection theory is straightforward, but since we have been unable to find these results in the literature in the form most convenient for our purpose, we give in an Appendix an exposition of this axiomatic "calculus of cycles." The proofs are essentially due to Weil [15], ch. VII, and Chevalley [2].

Notation and terminology. We fix an algebraically closed field k and a universal domain Ω containing k.

Our terminology is in general that of Lang [5]: all varieties are (absolutely) irreducible and abstract. Varieties will always be endowed with the Zariski Ω-topology, and the term *subvariety* will mean a closed irreducible subset in the sense of this topology. An everywhere defined rational map (or regular map) will be called a *morphism*; *isomorphisms* and *automorphisms* of varieties are defined accordingly.

We shall be dealing with certain categories of algebraic varieties. (For the notion of a category, see [10].) All varieties and morphisms in such a category $\boldsymbol{V}$ will be assumed to be defined over k. Unless otherwise specified, the term "generic" will mean "generic with respect to k."

Let $f\colon X \rightarrow Y$ be a morphism of algebraic varieties. We denote the graph of f by Γ_f. Let A' be a subvariety of X, A the closure of the set-theoretic image $f(A')$ in Y, and K a common field of definition for f and A'. There is a canonical injection $\hat{f}$ of $K(A)$ into $K(A')$, and A and A' have the same dimension if and only if the extension $K(A')/\hat{f}(K(A))$ is finite algebraic. When this is so, denote the degree of this extension by $[A'\colon A]$; otherwise set $[A'\colon A]=0$. When the degree $[A'\colon A]$ is finite, its separable part $[A'\colon A]_s$ is equal to the number of distinct points of A' lying above a generic point of A (over K).

If U is an irreducible locally closed subset of a variety X (so that U

is itself a variety), denote the canonical injection morphism by $i_U\colon U \to X$. Write $\pi_X\colon X \times Y \to X$ and $\pi_Y\colon X \times Y \to Y$ for the set-theoretic projections from a product onto its factors. If A is a subvariety of $X \times Y$, the *closed projection* of A on X is the closure of $\pi_X(A)$ in the Zariski topology.

The letters X, Y, V will always denote varieties; A, B, C, D (sometimes with primes or subscripts) will denote subvarieties of an ambient variety; Z, T (also with primes or subscripts) will denote cycles on an ambient variety. We denote the support of a cycle Z by $|Z|$.

Finally, Λ will denote a fixed integral domain of characteristic 0. In Section 1 and the Appendix, Λ is otherwise arbitrary; in Sections 2 and 3, Λ is assumed to contain the field $\boldsymbol{Q}$ of rational numbers. Λ will serve as coefficient-domain for the cycles on our varieties.

1. Intersection categories. Denote by $\boldsymbol{V}^*$ the category of all algebraic varieties defined over k, the morphisms in $\boldsymbol{V}^*$ being also defined over k. A category $\boldsymbol{V}$ of varieties and morphisms will be called *admissible* if it is a full subcategory[2] of $\boldsymbol{V}^*$ and satisfies the conditions:

i) if X, Y are in $\boldsymbol{V}$, so is $X \times Y$;

ii) if X' is isomorphic to an X in $\boldsymbol{V}$, the isomorphism being defined over k, then X' is in $\boldsymbol{V}$;

iii) if X is in $\boldsymbol{V}$ and U is an open subset of X, then U is in $\boldsymbol{V}$.

In order to speak of cycles on the varieties in an admissible category $\boldsymbol{V}$, we shall assume that we have, for each X in $\boldsymbol{V}$, a non-empty family $\mathfrak{S}(X)$ of "distinguished" subvarieties of X, subject to the following requirements:

iv) if $A \in \mathfrak{S}(X)$ and $A \subseteq A' \subseteq X$ then $A' \in \mathfrak{S}(X)$;

v) for any X, Y in V and any $A \subseteq X \times Y$ we have $A \in \mathfrak{S}(X \times Y)$ if and only if the closed projection of A on X (resp. on Y) is in $\mathfrak{S}(X)$ (resp. $\mathfrak{S}(Y)$);

vi) if X and X' in $\boldsymbol{V}$ are isomorphic, the isomorphism establishes a bijection between the sets $\mathfrak{S}(X)$ and $\mathfrak{S}(X')$. (The subvarieties $A \in \mathfrak{S}(X)$ need not be in $\boldsymbol{V}$; and if A is not in $\boldsymbol{V}$, A need not be defined over k.) By a *cycle* on a variety X in $\boldsymbol{V}$ we shall mean an element of the free Λ-module $\mathfrak{Z}(X)$ generated by the $A \in \mathfrak{S}(X)$. The module $\mathfrak{Z}(X)$ is graded by the dimensions of the $A \in \mathfrak{S}(X)$, and we shall deal only with homogeneous elements of $\mathfrak{Z}(X)$. For each $A \in \mathfrak{S}(X)$ we denote by $(A;-)$ the unique homomorphism of $\mathfrak{Z}(X)$

---

[2] A category C is a *full subcategory* of a category C^* if $C \subset C^*$ and if for any objects A, B in C the set of morphisms $A \to B$ in the category C is the same as the set of morphisms $A \to B$ in C^*.

into Λ such that $(A;A)=1$ and $(A;B)=0$ for all $B\in\mathfrak{S}(X)$, $B\neq A$. Every cycle Z on X then has a unique expression of the form $Z=\sum(A;Z)A$, summed over all $A\in\mathfrak{S}(X)$. The *components* of a cycle Z are those $A\in\mathfrak{S}(X)$ for which $(A;Z)\neq 0$. Since we assume Z homogeneous, all components of Z have the same dimension. (Taking $\mathfrak{S}(X)$ to be the family of simple subvarieties of X and Λ to be the ring $\boldsymbol{Z}$ of rational integers, we obtain the definition of cycle in the sense of Weil [15].)

Cartesian product of cycles is defined in the usual way (so as to be a bilinear operation). If $A\in\mathfrak{S}(X\times Y)$ we set $\mathrm{pr}_X A=[A:A_1]A_1$, where A_1 is the closed projection of A on X. Extend this operation linearly to cycles to obtain the *algebraic projection.* We also define the "inverse" of the algebraic projection by setting $\mathrm{pr}_X{}^{-1}B=B\times Y$ for any $B\in\mathfrak{S}(X)$, and extending linearly to cycles. The resulting operations pr_X, $\mathrm{pr}_X{}^{-1}$ are homomorphisms of Λ-modules.

Let $A_1,\cdots,A_s$ be subvarieties of a variety X in $\boldsymbol{V}$. If C is a component of $A_1\cap A_2\cap\cdots\cap A_s$, C will be called *proper* (on X) if $C\in\mathfrak{S}(X)$ and

$$\dim C=\sum_j\dim A_j-(s-1)\dim X.$$

If all components of $A_1\cap A_2\cap\cdots\cap A_s$ are proper on X (and in particular if the intersection of the A_j is empty), we say that $A_1,\cdots,A_s$ *intersect properly* on X. (Again, this reduces to the definition in [15] if C is simple on X.) Note that the existence of a proper component of $A_1\cap\cdots\cap A_s$ implies that all A_j are in the family $\mathfrak{S}(X)$.

Consider an ordered quadruple (X,A,B,C), where X is a variety in $\boldsymbol{V}$ and C is a proper component of $A\cap B$ on X. By an *intersection multiplicity* in the category $\boldsymbol{V}$ (with values in Λ) we shall mean a function which assigns to each such ordered quadruple a unique non-zero element of Λ. We denote this element by $(C;A\cdot B)_X$ and call it the *intersection multiplicity* of A and B at C on X. In general we suppress mention of the ambient variety, and write simply $(C;A\cdot B)$. Implicit in the existence of an intersection multiplicity on $\boldsymbol{V}$ is the existence of the families $\mathfrak{S}(X)$ of "distinguished" subvarieties, and the intersection multiplicity is only given relative to these $\mathfrak{S}(X)$.

If there is an intersection multiplicity in the category $\boldsymbol{V}$, we define the intersection-product of cycles in the usual fashion: if A and B intersect properly on X, set

$$A\cdot B=\sum(C;A\cdot B)C,$$

summed over all $C\in\mathfrak{S}(X)$. (If mention of the ambient variety is necessary, we write $(A\cdot B)_X$.) Extend this operation bilinearly to cycles. If $Z=\sum_i a_iA_i$ and $T=\sum_j b_jB_j$ are cycles on X, we set

$$(C;Z\cdot T)=\sum_{i,j}a_ib_j(C;A_i\cdot B_j),$$

summed over all A_i and B_j such that C is a proper component of $A_i\cap B_j$ on X. We use this notation even when the intersection-product $Z\cdot T$ is not defined on X, i.e., when $|Z|\cap|T|$ has improper components. (If $|Z|\cap|T|$ has no proper components, $(C;Z\cdot T)=0$ for all $C\in\mathfrak{S}(X)$.)

Now let $f\colon X\rightarrow Y$ be a morphism in V. Denote by $\mathfrak{Z}_f(X)$ the submodule of $\mathfrak{Z}(X)$ generated by those $A\in\mathfrak{S}(X)$ such that the closure of $f(A)$ is in $\mathfrak{S}(Y)$ and such that the intersection $\Gamma_f\cdot(A\times Y)$ is defined on $X\times Y$; and denote by $\mathfrak{Z}^f(Y)$ the submodule of $\mathfrak{Z}(Y)$ generated by those $B\in\mathfrak{S}(Y)$ for which $\Gamma_f\cdot(X\times B)$ is defined on $X\times Y$. The *direct image* of a cycle $Z\in\mathfrak{Z}_f(X)$ is defined by

$$f_*(Z)=\mathrm{pr}_Y[\Gamma_f\cdot(Z\times Y)],$$

and the *inverse image* of a cycle $T\in\mathfrak{Z}^f(Y)$ is given by

$$f^*(T)=\mathrm{pr}_X[\Gamma_f\cdot(X\times T)].$$

The maps $f_*\colon\mathfrak{Z}_f(X)\rightarrow\mathfrak{Z}(Y)$ and $f^*\colon\mathfrak{Z}^f(Y)\rightarrow\mathfrak{Z}(X)$ so defined are homomorphisms of Λ-modules. If the family $\mathfrak{S}(X)$ is the family of all subvarieties of X, then the submodule $\mathfrak{Z}_f(X)$ is all of $\mathfrak{Z}(X)$.

We now give the following definition:

Definition 1.1. *An intersection category* is an admissible category $\boldsymbol{V}$ of algebraic varieties and morphisms which satisfies the following six axioms:

(IC_1) ("Dimension theorem") If C is a component of an intersection $A\cap B$ on a variety X in $\boldsymbol{V}$, then $\dim C\geqq\dim A+\dim B-\dim X$.

(IC_2) There exists a commutative and associative intersection multiplicity in the category $\boldsymbol{V}$, relative to the families $\mathfrak{S}(X)$, where $\mathfrak{S}(X)$ is the family of *all* subvarieties of X.

(IC_3) For X in $\boldsymbol{V}$ and $A\subseteq X$ we have $(A\cdot X)_X=A$.

(IC_4) For X, Y in $\boldsymbol{V}$, A, $B\subseteq X$, and M, $N\subseteq Y$, we have

$$(A\times M)\cdot(B\times N)=A\cdot B\times M\cdot N$$

whenever one side is defined (the other side then being defined also.)

(IC_5) Let X, Y be varieties in $\boldsymbol{V}$. Let $A\subseteq X$ and $B\subseteq X\times Y$ be such that the restriction of π_X to B is proper (as a morphism of B into X) and the intersection $B\cdot(A\times Y)$ is defined on $X\times Y$. Then $(\mathrm{pr}_X B)\cdot A$ is defined on X, and

$$\mathrm{pr}_X[B\cdot(A\times Y)]=(\mathrm{pr}_X B)\cdot A,$$

(IC_6). Let $(C; A \cdot B)_X$ be defined on an X in $\boldsymbol{V}$, and let U be an open subset of X such that $U \cap C \neq \phi$. Then the intersection multiplicity $(U \cap C; (U \cap A) \cdot (U \cap B))_U$ is defined, and is equal to $(C; A \cdot B)_X$.

The commutative and associative properties of the intersection multiplicity are expressed by the relations $(C; Z \cdot T) = (C; T \cdot Z)$ and $(C; Z \cdot (T \cdot W)) = (C; (Z \cdot T) \cdot W)$ for all $C \subseteq X$ and all Z, T, $W \in \mathfrak{Z}(X)$.

Axioms (IC_4) and (IC_5) extend immediately to cycles, and will be used in this extended form. Axiom (IC_6) is not used in the development of the calculus of cycles, but is of crucial importance in the proof of Theorem 3.1. It is of course useful in passing from global to local results; see the remark following corollary A. 7 in the Appendix.

2. Quotients of algebraic varieties.

2. 1. Quotient structures on algebraic varieties. Let X be an algebraic variety, and denote by $\mathfrak{G}(X)$ the group of all automorphisms of X. For $x \in X$ and $\sigma \in \mathfrak{G}(X)$ we write x^σ instead of $\sigma(x)$, and for any subset $U \subseteq X$ we set $U^\sigma = \{x^\sigma : x \in U\}$. The group law in $\mathfrak{G}(X)$ is written as composition of mappings—$\sigma\tau = \sigma \circ \tau$—so that $U^{\sigma\tau} = (U^\tau)^\sigma$. The action of any finite subgroup $\mathfrak{g}$ of $\mathfrak{G}(X)$ defines an equivalence relation on the points of X. Let $X/\mathfrak{g}$ denote the quotient space, endowed with the quotient topology (of the Ω-topology on X), and let $\lambda : X \to X/\mathfrak{g}$ be the canonical projection. We give $X/\mathfrak{g}$ the structure of a ringed space by defining a subsheaf $\boldsymbol{O} = \boldsymbol{O}_{X/\mathfrak{g}}$ of the sheaf of germs of Ω-valued functions on $X/\mathfrak{g}$ as follows: the stalk $\boldsymbol{O}_y$ over a point $y \in X/\mathfrak{g}$ consists of those (germs of) functions ϕ defined on a neighborhood of y for which $\phi \circ \lambda$ is in the local ring of each of the points of $\lambda^{-1}(y)$ on X.

The ringed space $X/\mathfrak{g}$ need not be an algebraic variety. A necessary and sufficient condition that it be one is the following: every $\mathfrak{g}$-orbit $\mathfrak{g}x = \{x^\sigma : \sigma \in \mathfrak{g}\}$ is contained in an affine open set in X. (See [14], ch. III, no. 12, Proposition 19, and [3], Ch. V, Corollary 2 on p. 187; and cf. [13], § 1. 4.) In the sequel we shall always assume that this condition is satisfied: we consider only those pairs $(X, \mathfrak{g})$ for which the quotient space $X/\mathfrak{g}$ is an algebraic variety, with the structure given above.

This being the case, the sheaf $\boldsymbol{O}_{X/\mathfrak{g}}$ is the structure sheaf (= sheaf of local rings) of $X/\mathfrak{g}$; the projection map λ is a morphism, and the morphisms $f : X/\mathfrak{g} \to V$ of $X/\mathfrak{g}$ into another variety V are in one-to one correspondence with the morphisms $f' : X \to V$ which are constant on orbits. Moreover, λ is a proper morphism, and in fact a covering ([3], Ch. IV, Proposition 6 on p. 145 and Proposition 3 on p. 161; see also [13], 1. 1).

If all $\sigma \in \mathfrak{g}$ (considered as automorphisms of X) are defined over k we say that $\mathfrak{g}$ is defined over k. If in addition k is a field of definition for X, we say that the pair $(X, \mathfrak{g})$ is defined over k.

Let $(X, \mathfrak{g})$ be defined over k. The group $\mathfrak{g}$ acts on the field $k(X)$ by sending $\phi \to \phi^\sigma$, where $\phi^\sigma(x) = \phi(x^\sigma)$ for every $x \in X$ such that x^σ is in the domain of definition of ϕ. The quotient $X/\mathfrak{g}$ and the projection map are also defined over k, and the "adjoint map" λ is a canonical isomorphism of $k(X/\mathfrak{g})$ with the fixed field $k(X)^{\mathfrak{g}}$ of $\mathfrak{g}$ in $k(X)$. The extension $k(X)/\hat{\lambda}(k(X/\mathfrak{g}))$ is a finite Galois extension with group $\mathfrak{g}$, and therefore $\lambda\colon X \to X/\mathfrak{g}$ is a *Galois covering* (with group $\mathfrak{g}$). If X is an affine variety, so is $X/\mathfrak{g}$; and if X is normal, $X/\mathfrak{g}$ is also normal. If X is nonsingular *and if no* $\sigma \in \mathfrak{g}$ *except the identity has fixed points on* X, then $X/\mathfrak{g}$ will be nonsingular. If $(X, \mathfrak{g})$ and $(U, \mathfrak{h})$ are two pairs as above, the quotient $(X \times U)/(\mathfrak{g} \times \mathfrak{h})$ may be identified with the product $X/\mathfrak{g} \times U/\mathfrak{h}$. The proofs of these properties of quotients may be found in [12], Ch. III, no. 12, and [6], Proposition 2.

We note also the following two results. Let Y be a normal variety defined over the algebraically closed field k. For any finite Galois extension L of $k(Y)$, with group $\mathfrak{g}$, let X be the normalization of Y in the field L. Then the group $\mathfrak{g}$ acts as a group of k-automorphisms of X, and Y is k-isomorphic to $X/\mathfrak{g}$. ([14], Ch. VI, no. 7, and [3], Ch. V, Proposition 3 on p. 189.) Secondly, let X be a normal projective variety, and x a generic point of X over k. Let $\mathfrak{g}$ be a finite group of k-automorphisms of X, and let y be the Chow point of the "generic orbit" $\sum_{\sigma \in \mathfrak{g}} x^\sigma$. Let Y_0 be the locus of y over k, and Y the normalization of Y_0 in its function field. Then Y is k-isomorphic to $X/\mathfrak{g}$ (and hence the latter is projective), and the natural birational morphism $Y \to Y_0$ is bijective ([6], Proposition 3). That Y_0 itself need not be normal is shown by an example of Nagata [9].

Let $(X, \mathfrak{g})$ and $(X', \mathfrak{g}')$ be defined over k and let $\lambda\colon X \to X/\mathfrak{g}$ and $\lambda'\colon X' \to X'/\mathfrak{g}$ be the natural projections. If there exist a (surjective) isomirphism $i\colon X \to X'$ and a (surjective) isomorphism $j\colon \mathfrak{g} \to \mathfrak{g}'$ such that $i(x^\sigma) = i(x)^{j(\sigma)}$ for any $x \in X$ and $\sigma \in \mathfrak{g}$, we say that $(X, \mathfrak{g})$ and $(X', \mathfrak{g}')$ are *equivalent*. When this is so, i defines by passage to the quotient an isomorphism $i'\colon X/\mathfrak{g} \to X'/\mathfrak{g}'$ such that $i' \circ \lambda = \lambda'$. Let $\theta\colon X \to Y$ and $\theta'\colon X' \to Y$ be morphisms onto the same variety Y which are constant on orbits. Then θ (resp. θ') induces a morphism $\tau\colon X/\mathfrak{g} \to Y$ (resp. $\tau'\colon X'/\mathfrak{g}' \to Y$). We say that the triples $(X, \mathfrak{g}, \theta)$ and $(X', \mathfrak{g}', \theta')$ are *equivalent* if $(X, \mathfrak{g})$ and $(X', \mathfrak{g}')$ are equivalent and if $\tau = \tau' \circ i'$. For equivalent triples, τ is an isomorphism if and only if τ' is an isomorphism.

Definition 2.1. A *quotient structure* on the variety Y is an equivalence class of triples $(X, \mathfrak{g}, \theta)$ such that the morphism $\tau\colon X/\mathfrak{g} \to Y$ induced by θ is an isomorphism. We say that $\theta\colon X \to Y$ *endows* Y *with a quotient structure*; we call Y a *quotient variety* of X by $\mathfrak{g}$, and write $Y \leftarrow [X, \mathfrak{g}, \theta]$.

A single variety may be endowed with many different quotient structures. Also note that any variety X can be given a *trivial quotient structure* by taking $\mathfrak{g} = \{1\}$ and $\theta = \mathrm{id}_X$ (the identity automorphism of X).

Let $Y \leftarrow [X, \mathfrak{g}, \theta]$ with X normal. Then Y is also normal, and X is the normalization of Y in the field $k(X)$. We have already noted that the converse is true.

Note that if $Y \leftarrow [X, \mathfrak{g}, \theta]$ and $V \leftarrow [U, \mathfrak{h}, \phi]$ then

$$Y \times V \leftarrow [X \times U, \mathfrak{g} \times \mathfrak{h}, \theta \times \phi].$$

If $f\colon Y \to V$ is a morphism and $\rho\colon Y \to \Gamma_f$ is the inverse isomorphism of $\pi_Y \mid \Gamma_f\colon \Gamma_f \to Y$, then $\Gamma_f \leftarrow [X, \mathfrak{g}, \rho \circ \theta]$. Finally, if $W \subseteq Y$ is an open subset and $W' = \theta^{-1}(W)$, $\theta' = \theta \mid W'$, then $W \leftarrow [W', \mathfrak{g}, \theta']$.

2.2. Splitting groups and inertia groups. Let $Y \leftarrow [X, \mathfrak{g}, \theta]$. Subvarieties of X will be denoted by A', B', C', D', and their images on Y by A, B, C, D, respectively. For $A \subseteq Y$, a typical component of $\theta^{-1}(A)$ will be denoted by A'; A' will be said to *lie over* A. For each $A' \subseteq X$ we define two subgroups of $\mathfrak{g}$:

Definition 2.2. The *splitting group* of A' is the group

$$\mathfrak{g}^s(A') = \{\sigma \in \mathfrak{g} : A'^\sigma = A'\};$$

the *inertia group* of A' is the group

$$\mathfrak{g}^i(A') = \{\sigma \in \mathfrak{g} : x^\sigma = x \text{ for all } x \in A'\}.$$

If $\mathfrak{o}$ denotes the local ring of A' on X, the splitting (resp. inertia) group of A' is the same as the splitting (resp. inertia) group of $\mathfrak{o}$ in the sense of the Galois theory of local rings [1].

We associate with each $A \subseteq Y$ certain numerical characters:

Definition 2.3. $n(A) =$ number of components of $\theta^{-1}(A)$;
$d(A) = [A' : A] =$ degree of A' over A;
$d_s(A) = [A' : A]_s =$ separable degree of A' over A;
$d_i(A) = [A' : A]_i =$ inseparable degree of A' over A;
$l(A) = [\mathfrak{g}^i(A') : 1]/d_i(A)$.

Writing $g = [\mathfrak{g} : 1]$, we have the relation

$$n(A)d(A)l(A) = g = n(A)\,d_s(A)\,[\mathfrak{g}^i(A') : 1]$$

for any $A \subseteq Y$ and any $A' \subseteq X$ lying over A. Note that the components of $\theta^{-1}(A)$ are permuted transitively by the elements of $\mathfrak{g}$ ([3], Ch. V, Proposition 3 on p. 189), so that the above characters depend only on A and not on the choice of A'.

PROPOSITION 2.4. (i) $\mathfrak{g}^i(A')$ *is a normal subgroup of* $\mathfrak{g}^s(A')$.

(ii) *If* $B' \subseteq A'$ *then* $\mathfrak{g}^i(A') \subseteq \mathfrak{g}^i(B')$.

(iii) $n(A) = [\mathfrak{g} : \mathfrak{g}^s(A')]$ *and* $d_s(A) = [\mathfrak{g}^s(A') : \mathfrak{g}^i(A')]$.

(iv) *For any* $\sigma \in \mathfrak{g}$ *we have* $\mathfrak{g}^s(A'^{\sigma}) = \sigma\mathfrak{g}^s(A')\sigma^{-1}$, *and the same for* $\mathfrak{g}^i(A'^{\sigma})$.

Proof. The first two assertions are clear, as well as the statement about $n(A)$. For the rest: $d_s(A)$ is the separable degree of the morphism $\theta_A = \theta \mid A'$, hence is equal to the number of points of A' lying over a generic point of A; but this latter number is clearly equal to the number of $\sigma \in \mathfrak{g}^s(A')$ whose restrictions to A' are distinct, and this is the index $[\mathfrak{g}^s(A') : \mathfrak{g}^i(A')]$. And for any σ, τ in $\mathfrak{g}$ we have $A'^{\tau} = A'$ if and only if $(A'^{\sigma})^{\sigma\tau\sigma^{-1}} = A'^{\sigma}$, and the same for any point $x \in A'$. This proves the proposition.

It is not difficult to show that if the inertia group $\mathfrak{g}^i(A')$ is trivial, then $l(A) = 1$. If $\mathfrak{g}^i(A') \neq \{1\}$, $l(A)$—which is analogous to the ramification index for a valuation ring—need not be an integer; see the example in § 3.2.

2.3. Cycles on quotient varieties. Throughout the remainder of the paper Λ will denote a fixed integral domain containing the field of rational numbers.

Let $\boldsymbol{V}$ be an intersection category. We enlarge $\boldsymbol{V}$ to a category $\boldsymbol{V}'$ as follows: $\boldsymbol{V}'$ is the full subcategory of $\boldsymbol{V}^*$ whose objects are the varieties Y such that $Y \leftarrow [X, \mathfrak{g}, \theta]$ with X in $\boldsymbol{V}$ and with the pair $(X, \mathfrak{g})$ defined over k. Clearly $\boldsymbol{V}$ is a subcategory of $\boldsymbol{V}'$ and all varieties in $\boldsymbol{V}'$ are defined over k. For each $Y \in \boldsymbol{V}'$ we take $\mathfrak{S}(Y)$ to be the family of all subvarieties of Y; using this family and the domain Λ as coefficient domain we define cycles and the various operations on cycles as in § 1. The projection map $\theta : X \to Y$ induces mappings θ_*, θ^* on cycles, defined as follows. For any $A \subseteq Y$, let $A'_1, \cdots, A'_s$ be the distinct components of $\theta^{-1}(A)$, and set

$$A^* = \sum_j A'_j;$$

call A^* the *covering cycle* of A. Then define

$$\theta_s(A') = d(A)A \text{ and } \theta^s(A) = l(A)A^*,$$

and extend linearly to cycles. The resulting mappings $\theta_s\colon \mathfrak{Z}(X) \to \mathfrak{Z}(Y)$ and $\theta^s\colon \mathfrak{Z}(Y) \to \mathfrak{Z}(X)$ are homomorphics of Λ-modules, and θ^s is in fact easily seen to be injective. (Note that a cycle with rational integer coefficients on Y does not necessarily lift to a cycle with integer coefficients on X.)

For any cycle Z on Y we clearly have $\theta_s(\theta^s(Z)) = gZ$, where $g = [\mathfrak{g} : 1]$. This is a special case of the "projection formula" for θ (Proposition 3.3, (ii)), and the definition of θ^s was motivated by the necessity of having this relation.

From now until the end of § 3.1 we will be dealing with a fixed intersection category $\boldsymbol{V}$ and the category $\boldsymbol{V}'$ of quotients derived from it. Letters X, U will always denote varieties in $\boldsymbol{V}$ and Y, V will denote varieties in $\boldsymbol{V}'$; we shall use this convention without further reference. Let $Y \leftarrow [X, \mathfrak{g}, \theta]$. For $A \subseteq Y$ and any A' lying over A, the restriction $\theta_A = \theta \mid A'$ is a covering of A, so that $\dim A' = \dim A$ ([3], Ch. IV, Proposition 6 on p. 162). From this fact, and from the action of $\mathfrak{g}$ on X, we see easily that the dimension theorem (axiom (IC_1)) holds on all varieties in $\boldsymbol{V}'$, since by assumption it holds in $\boldsymbol{V}$.

The following lemmas on covering cycles will be used in the proof of Theorem 3.1. The second lemma is an immediate consequence of the first.

LEMMA 2.5. *Let C be a proper component of $A \cap B$ on Y. If C'_1 and C'_2 lie over C, then $(C'_1; A^* \cdot B^*)_X = (C'_2; A^* \cdot B^*)_X$, both multiplicities being defined.*

Proof. Let $C'_1{}^\sigma = C'_2$ for a $\sigma \in \mathfrak{g}$. If $\{A'_i\}$ (resp. $\{B'_j\}$) are the components of A^* (resp. B^*) containing C'_1, then C'_1 is a proper component of each intersection $A'_i \cap B'_j$; and the transforms of the A'_i and B'_j by σ are the components of A^* and B^* containing C'_2. We thus need only show that for any indices i, j we have $(C'_1; A'_i \cdot B'_j) = (C'_2; A'_i{}^\sigma \cdot B'_j{}^\sigma)$. This follows from the biregular invariance of the intersection multiplicity on X (see the remark following Corollary A.7 in the Appendix).

LEMMA 2.6. *If $A' \cdot B^*$ is defined on X, so is $A'^\sigma \cdot B^*$, for any $\sigma \in \mathfrak{g}$, and $\theta_s(A'^\sigma \cdot B^*) = \theta_s(A' \cdot B^*)$. As a result,*

$$\theta_s(A^* \cdot B^*) = n(A)\theta_s(A' \cdot B^*).$$

2.4. Products of quotient varieties. Let

$$Y \leftarrow [X, \mathfrak{g}, \theta] \text{ and } V \leftarrow [U, \mathfrak{h}, \phi],$$

and set $\mathfrak{G} = \mathfrak{g} \times \mathfrak{h}$ and $\Theta = \theta \times \phi$, so that $Y \times V \leftarrow [X \times U, \mathfrak{G}, \Theta]$. Denote subvarieties of U by M', N', and their images on V by M, N, with the same conventions as for subvarieties of X and Y.

LEMMA 2.7. (i) $(A \times M)^* = A^* \times M^*$.

(ii) $\mathfrak{G}^s(A' \times M') = \mathfrak{g}^s(A') \times \mathfrak{h}^s(M')$ *and*
$\mathfrak{G}^i(A' \times M') = \mathfrak{g}^i(A') \times \mathfrak{h}^i(M')$.

(iii) $d(A \times M) = d(A)d(M)$, *and the same for the other numerical characters of Definition* 2.3.

(iv) $\Theta_s = \theta_s \times \phi_s$ *and* $\Theta^s = \theta^s \times \phi^s$.

Proof. For $\sigma \in \mathfrak{g}$ and $\rho \in \mathfrak{h}$ we have $(A' \times M')^{\sigma \times \rho} = A'^{\sigma} \times M'^{\rho}$, which proves (i) and (ii). The relations $n(A \times M) = n(A)n(M)$ and $d_s(A \times M) = d_s(A)d_s(M)$ are also clear. To conclude the proof, we need only show that $d(A \times M) = d(A)d(M)$. Let K be a common field of definition for all varieties involved, and let x (resp. u) be a generic point of A' (resp. M') over K, such that x and u are independent over K. Then

$$\begin{aligned} d(A \times M) &= [K(x,u) : K(\theta(x), \phi(u))] \\ &= [K(x) : K(\theta(x))][K(u) : K(\phi(u))] \\ &= d(A)d(M), \end{aligned}$$

by [5], Ch. 3, Proposition 1, and [16], Lemma 1 on p. 114.

For a subvariety $W \subseteq Y \times V$ we denote by W_1 (resp. W_2) the closed projection of W on Y (resp. V); and similarly for $W' \subseteq X \times U$. Then we have:

LEMMA 2.8. *Let* $A \subseteq Y \times V$. *If* $A' \subseteq X \times U$ *lies over* A, *then* $(A')_1$ *lies over* A_1 *and* $(A')_2$ *lies over* A_2, *so that we may write* A'_1 *and* A'_2 *without ambiguity. Then:*

(i) $\mathfrak{G}^i(A') = \mathfrak{g}^i(A'_1) \times \mathfrak{h}^i(A'_2)$.

(ii) $\mathfrak{G}^s(A')$ *is a subgroup of* $\mathfrak{g}^s(A'_1) \times \mathfrak{h}^s(A'_2)$.

(iii) $d_s(A)$ *divides* $d_s(A_1)d_s(A_2)$ *and* $n(A_1)n(A_2)$ *divides* $n(A)$.

Proof. (i) and (ii) are immediate from the definitions, and (iii) follows from them.

It can also be shown that $d_i(A)$ divides $d_i(A_1)d_i(A_2)$, so that the same statement holds for the total degree $d(A)$.

That $\mathfrak{G}^s(A')$ may be a proper subgroup of $\mathfrak{g}^s(A'_1) \times \mathfrak{h}^s(A'_2)$ is shown by the following example: take $X = U$, $\mathfrak{g} \neq \{1\}$, and $\mathfrak{h} = \{1\}$. For the diagonal $\Delta \subset X \times X$ we have $\Delta_1 = \Delta_2 = X$; then $\mathfrak{G}^s(\Delta) = \{1\}$ but $\mathfrak{g}^s(X) = \mathfrak{g}$, $\mathfrak{h}^s(X) = \{1\}$.

The mappings θ_s, θ^s, Θ_s, Θ^s and the algebraic projections are related by the following formulas (in which $h = [\mathfrak{h} : 1]$):

Proposition 2.9. (i) $\mathrm{pr}_X \circ \Theta^s = h(\theta^s \circ \mathrm{pr}_Y)$.

(ii) $\Theta_s \circ \mathrm{pr}_X^{-1} = h(\mathrm{pr}_Y^{-1} \circ \theta_s)$.

(iii) $\mathrm{pr}_X^{-1} \circ \theta^s = \Theta^s \circ \mathrm{pr}_Y^{-1}$.

(iv) $\theta_s \circ \mathrm{pr}_X = \mathrm{pr}_Y \circ \Theta_s$.

Formulas (ii), (iii), and (iv) follow by brief computations from Lemma 2.7 and the definitions. The proof of (i) is more involvel, but since the result is not needed in this paper we shall omit the proof.

3. Intersection multiplicities on quotient varieties.

3.1. In this section we prove our principal result, in the following form:

Theorem 3.1. *Let V be an intersection category, and V' the category of quotient varieties derived from it. Then V' is an intersection category, with an intersection multiplicity which is compatible with the one given on V.*

We assume as in § 2.3 that the domain Λ in which the intersection multiplicity takes its values contains the field Q of rational numbers.

Let Y be a variety in V', $Y \leftarrow [X, \mathfrak{g}, \theta]$ for X in V. We shall define intersection multiplicities $(C; A \cdot B)_Y$ by using the morphism $\theta : X \rightarrow Y$ to carry over the multiplicities given on X (see Definition 3.2). In order for the quantities $(C; A \cdot B)_Y$ to be well-defined in the category V', they must be independent of the particular quotient structure used to define them. To prove that this is indeed the case, we shall work temporarily in a larger category than V': a category V'' in which we can distinguish between different quotient structures on the same underlying variety. We take, for each Y in V', a collection $\{Y_\alpha\}$ of copies of Y indexed by the distinct quotient structures on Y (Definition 2.1). If $(X, \mathfrak{g}, \theta)$ is a representative of the quotient structure α, we write $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$. In this way, each of the copies Y_α has associated with it a unique quotient structure. If $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$ and $V_\beta \leftarrow [U, \mathfrak{h}, \phi]$, we associate with the product $Y_\alpha \times V_\beta$ the unique quotient structure given by $Y_\alpha \times V_\beta \leftarrow [X \times U, \mathfrak{g} \times \mathfrak{h}, \theta \times \phi]$; similarly for products of several factors. A morphism $f : Y \rightarrow V$ in V' induces, for each copy Y_α of Y and V_β of V, a unique morphism $f_{\alpha\beta} : Y_\alpha \rightarrow V_\beta$. We associate with the graph $\Gamma_{\alpha\beta} \subset Y_\alpha \times V_\beta$ the unique quotient structure given by $\Gamma_{\alpha\beta} \leftarrow [X, g, \rho \circ \theta]$ where, as in § 2.1, ρ is the inverse isomorphism of $\pi_{Y_\alpha} \mid \Gamma_{\alpha\beta}\colon \Gamma_{\alpha\beta} \to Y_\alpha$. Let V'' be the category whose objects are the copies Y_α (as Y runs over V'), their products, and the graphs $\Gamma_{\alpha\beta}$;[3] the morphisms between two objects in V'' are their morphisms as algebraic varieties. We view V as a subcategory of V'' by identifying each X in V with the copy corresponding to the trivial quotient structure: $X \leftarrow [X, \{1\}, \mathrm{id}_X]$.

We define an intersection multiplicity in V''. Let $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$ and let C be a proper component of $A \cap B$ on Y_α—i. e., $\dim C = \dim A + \dim B - \dim Y_\alpha$. (Since the dimension theorem is valid in V', it holds also in V''.)

Definition 3.2. The *intersection multiplicity* of A and B at C on Y_α is

$$(C; A \cdot B)_{Y_\alpha} = \frac{l(A)l(B)}{l(C)} (C'; A^* \cdot B^*)_X,$$

where C' is any component of C^* and the multiplicity $(C'; A^* \cdot B^*)_X$ is that given in the category V. (The notation is that of §§ 2.2, 2.3.) The multiplicities on a product $Y_\alpha \times V_\beta$ are defined in the same way, using the covering $\theta \times \phi\colon X \times U \to Y_\alpha \times V_\beta$.

Because of the biregular invariance of intersection multiplicities in V (Corollary A.7), the quantity $(C; A \cdot B)_{Y_\alpha}$ is independent of the choice of representative $(X, \mathfrak{g}, \theta)$ for the quotient structure α. By Lemma 2.5 it is also independent of the choice of the component C' lying over C. Finally, the above-defined multiplicities are consistent with the given ones in the category V.

We use Definition 3.2 to obtain an intersection-product on cycles in the usual way. For a morphism $f\colon Y_\alpha \to V_\beta$ in V'' we define maps f_*, f^* on cycles as in Section 1. In proving Theorem 3.1, our main object will be to show that Corollary A.7 of the Appendix is valid in the category V''. Once we have this biregular invariance statement, we can conclude that the multiplicities given by Definition 3.2 are the same for all copies Y_α of Y; we thus obtain well-defined multiplicities on Y, independent of the quotient structure. Once this is achieved, we can discard the category V''.

The key result in the proof of Theorem 3.1 is the following proposition. It is worth noting that our definitions of θ_*, θ^*, and $(C; A \cdot B)_{Y_\alpha}$ are exactly those needed to make this proposition true.

PROPOSITION 3.3. *Let $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$ and yet Z, $T \in \mathfrak{Z}(Y_\alpha), Z' \in \mathfrak{Z}(X)$.*

(i) *If $Z \cdot T$ is defined on Y_α then $\theta^*(Z) \cdot \theta^*(T)$ is defined on X, and $\theta^*(Z \cdot T) = \theta^*(Z) \cdot {}^*\theta(T)$.*

---

[3] $Y_\alpha \times V_\beta$ is not one of the copies $(Y \times V)_\gamma$ of $Y \times V$, nor is $\Gamma_{\alpha\beta} \subset Y_\alpha \times V_\beta$ one of the copies $(\Gamma_f)_\delta$ of $\Gamma_f \subset Y \times V$.

(ii) ("Projection formula for θ.") *$Z'\cdot\theta^*(T)$ is defined on X if and only if $\theta_*(Z')\cdot T$ is defined on Y_α; when this is the case, $\theta_*(Z'\cdot\theta^*(T))=\theta_*(Z')\cdot T$.*

Proof. By linearity we may assume that Z, T, Z' are subvarieties A, B, A' (respectively).

(i) By hypothesis $A\cdot B$ is defined on Y_α; write $A\cdot B=\sum(C;A\cdot B)C$, summed over all components C. Then

$$\theta^*(A\cdot B)=\sum_C(C;A\cdot B)\theta^*(C)=\sum_C(C;A\cdot B)l(C)C^*$$

and

$$\begin{aligned}\theta^*(A)\cdot\theta^*(B)&=l(A)l(B)(A^*\cdot B^*)\\&=l(A)l(B)\sum_C(C';A^*\cdot B^*)C^*,\end{aligned}$$

by Lemma 2.5. Comparing coefficients of C^* in these two sums and using Definition 3.2, we are done.

(ii) The set-theoretic formula $\theta(A'\cap\theta^{-1}(B))=\theta(A')\cap B$, together with the dimension theorem ((IC$_1$)), proves the first assertion. Now set $A=\theta(A')$ and assume that $A\cdot B$ is defined; write $A\cdot B=\sum c_iC_i$. The coefficient of C_i in the cycle $\theta_*(A')\cdot B$ is

$$d(A)c_i=d(A)\frac{l(A)l(B)}{l(C_i)}(C_i';A^*\cdot B^*).$$

The cycle $\theta_*(A'\cdot\theta^*(B))=l(B)\theta_*(A'\cdot B^*)$ has the same components as $\theta_*(A')\cdot B$, and by Lemma 2.6 we have $n(A)\theta_*(A'\cdot B^*)=\theta_*(A^*\cdot B^*)$. In the cycle $A^*\cdot B^*$ each of the $n(C_i)$ components lying over C_i has the same coefficient $(C_i';A^*\cdot B^*)$ and the same degree $d(C_i)$ over its image. Hence the total coefficient of C_i in $\theta_*(A'\cdot\theta^*(B))$ is

$$l(B)\frac{1}{n(A)}n(C_i)d(C_i)(C_i';A^*\cdot B^*)=d(A)c_i,$$

which proves (ii).

The commutativity of the intersection-product on Y_α is clear from the definition. To prove the associativity of the intersection-product on Y_α:

$$\begin{aligned}gZ_1\cdot(Z_2\cdot Z_3)&=\theta_*\{\theta^*[Z_1\cdot(Z_2\cdot Z_3)]\}\\&=\theta_*\{\theta^*(Z_1)\cdot\theta^*(Z_2)\cdot\theta^*(Z_3)\}\\&=\theta_*\{\theta^*[(Z_1\cdot Z_2)\cdot Z_3]\}\\&=g(Z_1\cdot Z_2)\cdot Z_3\},\end{aligned}$$

when all intersections are defined; here, $g=[\mathfrak{g}:1]$. From Definition 3.2 it is also clear that $A\cdot Y_\alpha=A$ for any subvariety A, since the corresponding property holds on X. Thus axioms (IC_1), (IC_2) (in "global" form), and (IC_3) hold in the category $\boldsymbol{V}''$. The next two lemmas show that (IC_4) and (IC_5) are also valid. Let $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$ and $V_\beta \leftarrow [U, \mathfrak{h}, \phi]$, and set $\Theta = \theta \times \phi$.

LEMMA 3.5. *Let $A, B \subseteq Y_\alpha$ and $M, N \subseteq V_\beta$. Then $(A \times M) \cdot (B \times N)$ is defined on $Y_\alpha \times V_\beta$ if and only if $A \cdot B$ is defined on Y_α and $M \cdot N$ is defined on V_β; and when this is so,*

$$(A \times M) \cdot (B \times N) = A \cdot B \times M \cdot N.$$

Proof. Straightforward computation, using Propositions 3.3 and 2.9 to carry over the corresponding result on $X \times U$.

LEMMA 3.6. *Let $A \subseteq Y_\alpha$, and let $B \subseteq Y_\alpha \times V_\beta$ be such that $\pi_Y \mid B$ is proper. If $B \cdot (A \times V_\beta)$ is defined on $Y_\alpha \times V_\beta$, then,* $(\mathrm{pr}_{Y_\alpha} B) \cdot A$ *is defined on Y_α, and*

$$\mathrm{pr}_{Y_\alpha}[B \cdot (A \times V_\beta)] = (\mathrm{pr}_{Y_\alpha} B) \cdot A.$$

Proof. Since θ and Θ are coverings, the hypothesis on B implies that for any B' lying over B the restriction $\pi_X \mid B'$ is proper, and the intersection $B^* \cdot (A^* \times U)$ is defined on $X \times U$. By (IC_5) in the category V, we have

$$\mathrm{pr}_X[\Theta^*(B) \cdot (\theta^*(A) \times U)] = (\mathrm{pr}_X \Theta^*(B)) \cdot \theta^*(A).$$

Apply θ_* to both sides of this equality and use Proposition 2.9 and the projection formula for θ and for Θ; the desired result follows.

Conclusion of the proof of Theorem 3.1. Since we have axioms (IC_1) through (IC_5) in the category $\boldsymbol{V}''$, we know that Corollary A.7 of the Appendix is valid: any isomorphism in $\boldsymbol{V}''$ preserves intersection-products. Applying this to the isomorphism $i_{\alpha\beta}$: $Y_\alpha \to Y_\beta$ induced by the identity automorphism of Y, we deduce that intersection-products, when defined, are independent of quotient structure—i.e., $(C; A \cdot B)_{Y_\alpha}$ is independent of α. *provided* $A \cdot B$ is defined. Now suppose that $A \cdot B$ is not defined. Let C be a proper component of $A \cap B$ and F the union of the improper components. Let $Y_0 = Y - F$, $A_0 = A \cap Y_0$, $B_0 = B \cap Y_0$, $C_0 = C \cap Y_0$ (all non-empty sets). On Y_0 the intersection $A_0 \cdot B_0$ is defined, so that $(C_0; A_0 \cdot B_0)_{Y_0}$ is independent of the quotient structure on Y_0. If $Y_\alpha \leftarrow [X, \mathfrak{g}, \theta]$, we set $X_0 = \theta^{-1}(Y_0)$; X_0 is open in X and is $\mathfrak{g}$-invariant, so that $Y_0 \leftarrow [X_0, \mathfrak{g}, \theta \mid X_0]$. Using this quotient structure to compute $(C_0; A_0 \cdot B_0)_{Y_0}$ and using axiom (IC_6) on X, we find that $(C_0; A_0 \cdot B_0)_{Y_0} = (C; A \cdot B)_{Y_\alpha}$. Thus the latter quantity is independent of the quotient structure used to define it; denote it by $(C; A \cdot B)_Y$. We thus obtain a well-defined intersection-multiplicity on the category $\boldsymbol{V}'$.

The same argument as the above, now applied to an arbitrary open subset Y_0, shows that axiom (IC_6) holds in $\boldsymbol{V}'$. Applying (IC_6) to the "global" associativity formula $Z_1\cdot(Z_2\cdot Z_3)=(Z_1\cdot Z_2)\cdot Z_3$ yields the associativity of the symbol $(C;A\cdot B)_Y$, thus proving (IC_2). Finally, Lemmas 3.5 and 3.6 clearly carry over to $\boldsymbol{V}'$, so that axioms (IC_4) and (IC_5) are valid. We have already noted that (IC_1) and (IC_3) hold. Hence $\boldsymbol{V}'$ is an intersection category.

The proof of Theorem 3.1 is now complete. We shall apply it to the case when $\boldsymbol{V}$ is the category of all nonsingular varieties.

As was remarked in § 2.1, the quotient structures on a normal variety Y are in one-to-one correspondence with the normalizations of Y in Galois extensions of its function field. Suppose that a variety Y satisfies the following condition:

(N_n) *Y is normal, and there exists a finite tower of fields*

$$K_n\supseteq K_{n-1}\supseteq\cdots\supseteq K_1\supseteq K_0=k(Y)$$

such that each extension K_j/K_{j-1} is a finite Galois extension and such that the normalization of Y in K_n is nonsingular.

Denote by X_j the normalization of Y in K_j ($X_0=Y$) and by $\mathfrak{g}_j$ the Galois group of K_j/K_{j-1}. The above remark, together with the transitivity of the normalization process, implies that there exist morphisms $\theta_j\colon X_j\to X_{j-1}$ such that $X_{j-1}\leftarrow[X_j,\mathfrak{g}_j,\theta_j]$ for $j=1,\cdots,n$. Let $\boldsymbol{V}_n$ be the category of all varieties satisfying condition (N_n). Then $\boldsymbol{V}_0$ is just the category of all nonsingular varieties, and for each n, $\boldsymbol{V}_{n+1}$ is the category of quotient varieties derived from $\boldsymbol{V}_n$ (§ 2.3). Using Theorem 3.1 to perform an induction on n, we obtain the following result.

THEOREM 3.7. *Let $\boldsymbol{Q}$ be the category of all normal varieties Y having a nonsingular normalization in some field K obtainable from $k(Y)$ by a finite sequence of (finite) Galois extensions. Then $\boldsymbol{Q}$ is an intersection category, with an intersection multiplicity which agrees on nonsingular varieties with the usual one.*

We remark the following results explicitly: if Y is a variety in $\boldsymbol{Q}$ and if a proper component C of $A\cap B$ on Y is *simple* on Y, then $(C;A\cdot B)_Y$ agrees with the usual intersection multiplicity as defined in [15]. For we may remove the singular locus of Y to obtain a nonsingular variety Y_0, and the intersection multiplicitl defined on Y_0 by any quotient structure must agree with that defined by the trivial quotient structure.

We return for a moment to an arbitrary intersection category $\boldsymbol{V}$ and its derived category $\boldsymbol{V}'$ of quotients. If $Y\leftarrow[X,\mathfrak{g},\theta]$ we have

$$\theta_*(Z') = \mathrm{pr}_Y[\Gamma_\theta \cdot (Z' \times Y)]$$

and

$$\theta^*(Z) = \mathrm{pr}_X[\Gamma_\theta \cdot (X \times Z)]$$

for any $Z' \in \mathfrak{Z}(X)$, $Z \in \mathfrak{Z}(Y)$. (The first formula follows from Lemma A. 2 and the second is easily checked using Theorem A. 5.) The notation θ_*, θ^* is thus consistent with that introduced in Section 1 for arbitrary morphisms in an intersection category. The above expression for θ^* also gives an interpretation of the numerical character $l(A)$ as an intersection multiplicity:

$$l(A) = (A''; \Gamma_\theta \cdot (X \times A)),$$

where A'' is any component of $\Gamma_\theta \cap (X \times A)$.

3. 2. Examples. We first give an example to show that the numerical character $l(A)$ (Definition 2. 3) need not be an integer. Let V be the affine plane over a universal domain Ω of characteric $p > 0$, and set $X = V \times V \times \cdots \times V$ (p factors). The symmetric group $\mathfrak{S}_p$ on p letters operates on X by permutation of factors, each permutation being an automorphism of X. The quotient $Y = V(p)$ is the p-fold symmetric product of V, and the projection map $\theta\colon X \rightarrow Y$ is an Galois covering of degree $p!$. Let $\Delta' \subset X$ be the diagonal, and Δ its image in Y. Then Δ' is isomorphic to V and so has function field $k(u, v)$ over any perfect field $k \subset \Omega$, u and v being independent transcendentals over k. The image Δ has function field $k(u^p, v^p)$ by [14], Proposition 21 on p. 62. Since the inertia group of Δ' is all of $\mathfrak{S}_p$ and since $[k(u, v) : k(u^p, v^p)] = p^2$, we have

$$l(\Delta) = \frac{p!}{p^2} = \frac{(p-1)!}{p}.$$

(This example was previously noted in [7].) It can be shown that if we consider instead of the symmetric product the cyclic product $V_c(p) \leftarrow [X, \mathfrak{c}_p, \phi]$ (where $\mathfrak{c}_p$ is the group of cyclic permutations on p letters), then $[\Delta' : \phi(\Delta')] = p^2$, so that $l(\phi(\Delta')) = 1/p$. In this case the degree $[\Delta' : \phi(\Delta')]$ is *greater than* the degree $[X : V_c(p)]$ of the covering.

In view of the remark at the end of § 3. 1, we have an example of an intersection multiplicity $l(\Delta) = (\Delta''; \Gamma_\theta \cdot (X \times \Delta))$ whose fractional nature stems from inseparability phenomena in non-zero characteristic. However the following example shows that even in characteristic 0, when $l(A)$ is always an integer, fractional multiplicities may occur.

Let Ω have characteristic 0, and as above let V be the affine plane over Ω. Let $X = V \times V$ and $Y = V(2)$. We identify X with affine 4-space over Ω and Y with the Chow variety of positive 0-cycles of degree 2 on V (see [9]).

If (x_1, y_1, x_2, y_2) is a generic point for X over an algebraically closed field $k \subset \Omega$, we may take $(x_1 + x_2, x_1x_2, y_1 + y_2, y_1y_2, x_1y_2 + x_2y_1)$ as a generic point for Y over k. In this way we represent Y as a quartic hypersurface in affine 5-space, with equation

$$(2T_5 - T_1T_3)^2 - (T_1^2 - 4T_2)(T_3^2 - 4T_4) = 0.$$

The "diagonal" Δ in Y is the locus of the point $(2x_1, x_1^2, 2y_1, y_1^2, 2x_1y_1)$ over k. Let t, u, v, w be independent transcendentals over k and set

$A =$ locus of the point $P_1 = (t, \frac{1}{4}t^2, u, v, \frac{1}{2}tu)$ over k,
$B =$ locus of the point $P_2 = (t, w, u, \frac{1}{4}u^2, \frac{1}{2}tu)$ over k;

also we have

$\Delta =$ locus of the point $Q = (t, \frac{1}{4}t^2, u, \frac{1}{4}u^2, \frac{1}{2}tu)$ over k.

Clearly $\Delta = A \cap B$; since $\dim A = \dim B = 3$ and $\dim \Delta = 2$, Δ is a proper component of $A \cap B$ on Y and $(\Delta; A \cdot B)_Y$ is defined. We show that $A \cdot B = \frac{1}{2}\Delta$.

Let $\theta\colon X \to Y$ be the natural projection of $V \times V$ onto $V(2)$. If $A' \subset X$ is the locus over k of the point

$$P_1' = (\tfrac{1}{2}t, \tfrac{1}{2}u + \sqrt{\tfrac{1}{4}u^2 - v}, \tfrac{1}{2}t, \tfrac{1}{2}\mathrm{u} - \sqrt{\tfrac{1}{4}u^2 - v}),$$

we have clearly $\theta(P_1') = P_1$ and $\theta(A') = A$. Since $[k(P_1') : k(P)] = 2$ we conclude that $d(A) = 2$ and $n(A) = l(A) = 1$. In the same way, letting $B' \subset X$ be the locus over k of the point

$$P_2' = (\tfrac{1}{2}t + \sqrt{\tfrac{1}{4}t^2 - w}, \tfrac{1}{2}u, \tfrac{1}{2}t - \sqrt{\tfrac{1}{4}t^2 - w}, \tfrac{1}{2}u),$$

we find that $d(B) = 2$ and $n(B) = l(B) = 1$. It is easily checked that $A' \cap B' = \Delta'$, the diagonal in $X = V \times V$, and that $(\Delta'; A' \cdot B')_X = 1$. Then

$$(\Delta; A \cdot B)_Y = \frac{1}{l(\Delta)}(\Delta'; A^* \cdot B^*)_X = \tfrac{1}{2}(\Delta'; A' \cdot B')_X = \tfrac{1}{2},$$

as claimed.

Geometrically the subvarieties A and B arise in the following way. If we cut Y by a linear 3-space with equations $T_1 = t$, $T_3 = u$, we obtain as cross-section a quadratic cone in $T_2T_4T_5$-space having equation

$$(2T_5 - tu)^2 - (t^2 - 4T_2)(u^2 - 4T_4) = 0.$$

This cone has vertex $Q = (t, \frac{1}{4}t^2, u, \frac{1}{4}u^2, \frac{1}{2}tu)$, and we may specify two generators by setting

$$L_1 = \text{locus of } P_1 = (t, \tfrac{1}{4}t^2, u, v, \tfrac{1}{2}tu) \text{ over } k(t, u),$$
$$L_2 = \text{locus of } P_2 = (t, w, u, \tfrac{1}{4}u^2, \tfrac{1}{2}tu) \text{ over } k(t, u).$$

The generators L_1 and L_2 intersect at the vertex Q with multiplicity $\frac{1}{2}$: see [11], page 99. As (t, u) varies in the T_1T_3-plane, the vertex Q traces out the surface Δ and the generators L_1, L_2 sweep out the ruled threefolds A, B.

Appendix: The calculus of cycles.

In this appendix we prove the basic statments in the global intersection theory of cycles. We use the definitions and notation of Section 1. In particular, we denote by Λ an integral domain of characteristic 0; all cycle coefficients and intersection multiplicities are assumed to belong to Λ.

Let $\mathbf{V}$ be an intersection category. X and Y will denote varieties in $\mathbf{V}$ and $f\colon X \to Y$ will denote a morphism in $\mathbf{V}$. Recall that the graph Γ_f of f, being isomorphic to X, is in $\mathbf{V}$. The notation $W \subseteq X$ will always mean that W is a closed (irreducible) subvariety of X.

We begin with some elementary lemmas; the proofs are easy applications of (IC_3) and (IC_5), and are omitted.

LEMMA A.1. *Let* $Z \in \mathfrak{Z}(X)$ *and* $Z' \in \mathfrak{Z}(\Gamma_f)$. *The relations* $\mathrm{pr}_X Z' = Z$ *and* $Z' = \Gamma_f \cdot (Z \times Y)$ *are equivalent, and establish a* Λ*-isomorphism between* $\mathfrak{Z}(X)$ *and* $\mathfrak{Z}(\Gamma_f)$.

LEMMA A.2. *For* $A \subseteq X$ *let* A'' *be the closure of* $f(A)$ *in* Y. *Then* $f_*(A) = [A : A'']A''$.

We thus have an alternate expression for f_* which does not involve any intersection theory. In particular we see that the algebraic projection pr_X is equal to $(\pi_X)_*$. The corresponding formula for the "inverse projection," $\mathrm{pr}_X^{-1} = (\pi_X)^*$, is an easy consequence of Theorem A.5 below.

Let $W \subseteq Y$ be a subvariety such that W is in $\mathbf{V}$. The injection i_W is proper since W is closed in Y, and it is clear from Lemma A.2 that $(i_W)_*$ is a Λ-isomorphism of $\mathfrak{Z}(W)$ into $\mathfrak{Z}(Y)$, the image consisting of those cycles on Y carried by W. We shall usually identify $\mathfrak{Z}(W)$ with its image in $\mathfrak{Z}(Y)$. We also note the following: if $g\colon Y \to V$ is a morphism in $\mathbf{V}$ then for any cycle $Z \in \mathfrak{Z}(Y)$ carried by W we have $g_*(Z) = (g \mid W)_*(Z)$.

LEMMA A.3. *Let* f *be an isomorphism of* X *onto a subvariety of* Y. *Then the homomorphism* $f_*\colon \mathfrak{Z}(X) \to \mathfrak{Z}(Y)$ *is injective. If* $f(X) = Y$ *then* f_* *and* f^* *are bijections and are inverses of one another.*

LEMMA A.4. *Let* $W \subseteq X$ *with* W *in* $\mathbf{V}$. *For any* $A \subseteq X$, *the cycle*

$i_W{}^*(A)$ is defined if and only if W and A intersect properly on X; when this is so, $i_W{}^*(A) = (W \cdot A)_X$.

We are now able to prove one of the most important results in the calculus of cycles, the "projection formula." It is a generalization of (IC_5).

THEOREM A. 5. *Let $Z \in \mathfrak{Z}(X)$ be such that the restriction of f to each component of Z is proper, and let $T \in \mathfrak{Z}^f(Y)$. If $Z \cdot f^*(T)$ is defined on X, then $f_*(Z) \cdot T$ is defined on Y, and*

$$f_*(Z \cdot f^*(T)) = f_*(Z) \cdot T.$$

Proof. By linearity, we need only prove the theorem for subvarieties $Z = A$, $T = B$. The set-theoretic equality $f(A \cap f^{-1}(B)) = f(A) \cap B$ is clear. By hypothesis the intersections $\Gamma_f \cdot (X \times B)$ and $A \cdot f^*(B)$ are defined (on $X \times Y$ and X, respectively). From Lemma A. 1 we know that for any cycle $Z_0 \in \mathfrak{Z}(\Gamma_f)$ we have the relation $Z_0 = \Gamma_f \cdot [(\mathrm{pr}_X Z_0) \times Y]$. Applying this to the cycle $Z_0 = \Gamma_f \cdot (X \times B)$ we obtain

$$\Gamma_f \cdot (X \times B) = \Gamma_f \cdot [\mathrm{pr}_X(\Gamma_f \cdot X \times B)) \times Y] = \Gamma_f \cdot [f^*(B) \times Y],$$

all intersections being defined. Now we have

$$\begin{aligned} f_*(A \cdot f^*(B)) &= \mathrm{pr}_Y[\Gamma_f \cdot (A \cdot f^*(B) \times Y)] \\ &= \mathrm{pr}_Y[\Gamma_f \cdot (f^*(B) \times Y) \cdot (A \times Y)] \\ &= \mathrm{pr}_Y[\Gamma_f \cdot (X \times B) \cdot (A \times Y)] \\ &= \mathrm{pr}_Y[\Gamma_f \cdot (A \times Y)] \cdot B \\ &= f_*(A) \cdot B, \end{aligned}$$

by (IC_2), (IC_4), and (IC_5). This is the desired result.

We note the following: if f is proper and surjective, then $f_*(f^*(T)) = [X : Y]T$ for any $T \in \mathfrak{Z}^f(Y)$.

We now give two important corollaries of the projection formula. The first of these might be called the "relativization formula."

COROLLARY A. 6. *Let $W \subseteq X$, with W in V. Let $T \in \mathfrak{Z}(X)$ and $Z \in \mathfrak{Z}(W)$. If $(W \cdot T)_X$ is defined and if, as a cycle on W, it intersects Z properly on W, then $(Z \cdot T)_X$ is defined and*

$$(Z \cdot (W \cdot T)_X)_W = (Z \cdot T)_X.$$

The same result holds under the alternative hypothesis that $(W \cdot T)_X$ and $(Z \cdot T)_X$ are both defined.

The next corollary expresses the biregular invariance of the intersection multiplicity in V.

COROLLARY A. 7. *If $f\colon X \to Y$ is a (surjective) isomorphism and if Z, Z' are cycles on X such that $(Z\cdot Z')_X$ is defined, then*

$$f_*(Z\cdot Z') = f_*(Z)\cdot f_*(Z').$$

Proof. By Lemma A. 3 we have $f^*(f_*(Z')) = Z'$; then the projection formula yields $f_*(Z\cdot Z') = f_*(Z\cdot f^*(f_*(Z'))) = f_*(Z)\cdot f_*(Z')$, which proves the corollary.

The (stronger) local statement—the biregular invariance of $(C; Z\cdot Z')$ even when $Z\cdot Z'$ is not defined—is an immediate consequence of Corollary A. 7 and (IC_0).

THEOREM A. 8. *If Z, Z' are cycles in $\mathfrak{Z}^f(Y)$ such that $(Z\cdot Z')_Y$ is defined and is also in $\mathfrak{Z}^f(Y)$, then $f^*(Z)\cdot f^*(Z')$ is defined on X and*

$$f^*(Z\cdot Z') = f^*(Z)\cdot f^*(Z').$$

Proof. By (IC_4) we have

$$f^*(Z\cdot Z') = \mathrm{pr}_X[\Gamma_f\cdot (X\times Z\cdot Z')] = \mathrm{pr}_X[\Gamma_f\cdot (X\times Z)\cdot (X\times Z')],$$

and by the "relativization formula" (Corollary A. 6) we have

$$\Gamma_f\cdot (X\times Z)\cdot (X\times Z') = \{[\Gamma_f\cdot (X\times Z)]_{X\times Y}\cdot [\Gamma_f\cdot (X\times Z')]_{X\times Y}\}_{\Gamma_f}.$$

If we set $\phi = \pi_X \mid \Gamma_f$ and $W = \Gamma_f\cdot (X\times Z)$, $W' = \Gamma_f\cdot (X\times Z')$, we have

$$f^*(Z\cdot Z') = \phi_*((W\cdot W')_{\Gamma_f}),$$

and the theorem follows from Corollary A. 7.

We now state (without proof) two further results which complete the list of basic formulas in the calculus of cycles. Let $f\colon X\to Y$, $g\colon U\to V$, and $h\colon Y\to W$ be morphisms in V; then we have morphisms $h\circ f\colon X\to W$ and $f\times g\colon X\times U\to Y\times V$.

PROPOSITION A. 9. (i) *$Z\times T\in \mathfrak{Z}^{f\times g}(Y\times V)$ if and only if $Z\in\mathfrak{Z}^f(Y)$ and $T\in\mathfrak{Z}^g(V)$; when this is so,*

$$(f\times g)^*(Z\times T) = f^*(Z)\times g^*(T).$$

(ii) *For any $Z\in\mathfrak{Z}(X)$ and $T\in\mathfrak{Z}(U)$ we have*

$$(f\times g)_*(Z\times T) = f_*(Z)\times g_*(T).$$

PROPOSITION A. 10. *Let $T\in\mathfrak{Z}(W)$. When either side is defined, we have $(h\circ f)^*(T) = f^*(h^*(T))$. If f is a closed map, we hav also $(h\circ f)_*(Z) = h_*(f_*(Z))$ for any $Z\in\mathfrak{Z}(X)$.*

The assertions regarding direct images are clear, using Lemma A. 2;

those regarding inverse images are proved by straightforward calculations. For a slightly more precise result than Proposition A. 10, see [2], corollary to Proposition 7. The proof given there can be carried over without difficulty to any intersection category.

Many other useful results are derivable from the formalism we have set up. For example, the definition and properties of algebraic families of cycles given in [2] can be carried over almost without change to any intersection category. Also, equivalence relations on cycles can be studied in this axiomatic framework, using Samuel's notion of an adequate equivalence relation [12].

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] S. Abhyankar, *Ramification Theoretic Methods in Algebraic Geometry*, Annals of Mathematics Studies, No. 43, Princeton, 1959.

[2] C. Chevalley, "Les classes d'équivalence rationelle (I)," *Séminaire C. Chevalley*, 2e année: 1958 ("*Anneaux de Chow et Applications*"), expose 2.

[3] ———, *Fondements de la Géométrie Algébrique*, Secrétariat Mathématique (Paris), mimeographed.

[4] Grothendieck, A., "Sur quelques propriétés fondamentales en théorie des intersections," *Séminaire C. Chevalley*, 2e année: 1958, exposé 4.

[5] S. Lang, *Introduction to Algebraic Geometry*, Interscience (New York), 1958.

[6] ———, "Unramified class field theory over function fields in several variables," *Annals of Mathematics*, vol. 64 (1956), p. 285.

[7] ——— and J.-P. Serre, "Errata," *American Journal of Mathematics*, vol. 81 (1959), p. 279.

[8] M. Nagata, "On the normality of the Chow variety of positive 0-cycles of degree m in an algebraic variety," *Memoirs of the College of Science, University of Kyoto, Ser.* A (*Mathematics*), vol. 29 (1955), p. 165.

[9] D. G. Northcott, *An Introduction to Homological Algebra*, Cambridge University Press, 1960.

[10] P. Samuel, *Méthodes d'algèbre abstraite en géométrie algébrique*, Springer (Berlin), 1955.

[11] ———, "Relations d'équivalence en géométrie algébrique," *Proceedings of the International Congress of Mathematicians* (1958), Cambridge University Press, 1960, p. 470.

[12] J.-P. Serre, "Espaces fibrés algébriques," *Séminaire C. Chevalley*, 2e année: 1958, exposé 1.

[13] ———, *Groupes algébriques et corps de classes*, Actualités Scientifiques et Industrielles no. 1264, Hermann (Paris), 1959.

[14] A. Weil, *Foundations of Algebraic Geometry*, American Mathematical Society Colloquium Publications, No. XXIX (New York), 1946.

[15] O. Zariski and P. Samuel, *Commutative Algebra*, vol. I, D. van Nostrand Co. (Princeton), 1958.

AUTOTOPISM GROUPS OF SOME FINITE NON-ASSOCIATIVE ALGEBRAS.*[1]

By REUBEN SANDLER.

---

1. Introduction. Many classes of finite non-associative algebras have been studied and their autotopism groups determined. From the point of view of projective geometry, the algebras of interest are the algebras which are also *division* algebras, since every such algebra coordinatizes a finite projective plane and furthermore, for a plane coordinatized by a division algebra, the autotopism group of the algebra is isomorphic with the collineation group of the plane modulo the normal subgroup generated by the translations and shears (see [5] for a more complete exposition of the above).

In this paper, I propose to define a generalization of a class of algebras discussed by Hughes and Kleinfeld [3] and [4], and Albert [1], and study the autotopism groups of these algebras. Some of these algebras are shown to be division algebras and coordinatize new finite projective planes. Finally, the question of isotopy between the division algebras (equivalent to the problem of isomorphism between the planes they coordinatize) is discussed.

2. Preliminary results. Let K be the finite field of degree m over the field $F = GF(q) = GF(p^r)$. Then K has degree mr over the prime field $GF(p)$, and is cyclic over $GF(p)$. The automorphism T defined by

$$(1) \qquad xT = x^p$$

for x in K generates the automorphism group of K over $GF(p)$. K is also cyclic over F, and its automorphism group over F is generated by $S = T^r$, so

$$(2) \qquad xS = x^{p^r} = x^q$$

for x in K. For x, y in K, we write

$$xy = xR(y) = yL(x).$$

We now define the algebra $\Omega_{\delta,n,S}$ of degree n over K and degree mn over F, for every δ in K, and for any $n \leq m$ as follows: $\Omega_{\delta,n,S}$ is an n-dimensional

---

* Received July 19, 1961.

[1] This research is from the author's doctoral dissertation at the University of Chicago, September, 1961.

vector space over K with basis $1, \lambda, \lambda^2, \cdots, \lambda^{n-1}$, and with multiplication defined by

$$xλ^i = \lambda^i (xS^i) \tag{3}$$

for x in K and $i < n$,

$$(\lambda^i a_1)(\lambda^j a_2) = \lambda^i \lambda^j (a_1 S^j) a_2 \text{ for } a_1, a_2 \text{ in } K, \tag{4}$$

$$\lambda^i \lambda^j = \begin{cases} \lambda^{i+j} \text{ for } i+j<n \\ \delta \text{ for } i+j=n \\ \lambda^{i+j-n}\delta \text{ for } i+j>n \end{cases} \tag{5}$$

We also assume both distributive laws.

A typical element of $\Omega_{\delta,n,S}$, then, will be of the form

$$x_0 + \lambda x_1 + \cdots + \lambda^{n-1} x_{n-1} \tag{6}$$

and equations (3), (4), and (5) tell us that we can write:

$$L_{x_0+\lambda x_1+\cdots+\lambda^{n-1}x_{n-1}} \tag{7}$$

$$= \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_{n-1} \\ \delta(x_{n-1}S) & (x_0 S) & (x_1 S) & \cdots & (x_{n-2}S) \\ \delta(x_{n-2}S^2) & \delta(x_{n-1}S^2) & (x_0 S^2) & \cdots & (x_{n-3}S^2) \\ \vdots & \vdots & \vdots & & \vdots \\ \delta(x_1 S^{n-1}) & \delta(x_2 S^{n-1}) & \delta(x_3 S^{n-1}) & \cdots & (x_0 S^{n-1}) \end{bmatrix}$$

where L_z is the matrix for left multiplication by the element z of $\Omega_{\delta,n,S}$. We could also write down a matrix for right multiplication by z, R_z, but we will not need that during the course of this paper. In order to differentiate between multiplication in the field K and the algebra $\Omega_{\delta,n,S}$, define

$$x \circ y = xR_y = yL_x \tag{8}$$

for x and y in $\Omega_{\delta,n,S}$.

Hughes and Kleinfeld [4] have defined a more general class of these algebras in the case $n = 2$, and Hughes [3] has shown that the autotopism groups of their algebras are all solvable. Albert [1] has actually computed the autotopism groups in the case $n = 2$, $m > 2$. In this paper, we will exhibit the groups for the case $n = m = 2$ and then study the case $n > 2$.

As usual, define the right nucleus, N_r, of Ω to be the set of elements a in Ω such that

$$x \circ (y \circ a) = (x \circ y) \circ a \tag{9}$$

for all x, y in Ω. The *left* and *middle nuclei* N_l and N_m are analogously defined. The *center* of Ω is the set of all z in Ω in the intersection of all three nuclei and such that

$$z \circ x = x \circ z \tag{10}$$

for all x in Ω.

We can begin with the following:

THEOREM 1. *The right nucleus of $\Omega_{\delta,n,S}$ is equal to its middle nucleus, and both consist of all elements of the form*

$$x = x_0 + \lambda \cdot 0 + \cdots + \lambda^{n-1} \cdot 0, \tag{11}$$

unless $n = m$ and δ lies in a proper subfield of K. In the latter case, the two nuclei are still equal, but consist of all elements

$$x = x_0 + \lambda^s x_s + \lambda^{2s} x_{2s} + \cdots + \lambda^{n-s} x_{n-s} \tag{12}$$

where the x_i are in K. The center of Ω, $Z(\Omega)$, consists of all elements of the form (11) *where $x_0 S = x_0$, i.e., $Z(\Omega) \cong F$.*

Proof. As a typical case, we give the proof for the middle nucleus. In determining the nuclei, we note that the distributive laws imply that it suffices to satisfy the definitions with the elements x and y being restricted to elements of the form $\lambda^i x_i$, instead of all x, y in Ω. For if

$$(\lambda^i x_i) \circ [(\lambda^j y_j) \circ a] = [(\lambda^i x_i) \circ (\lambda^j y_j)] \circ a \qquad \text{for all } i, j, x_i, y_j,$$

then, clearly, a will satisfy (9). We observe first that any element of the form $x_0 + \lambda \cdot 0 + \lambda^2 \cdot 0 + \cdots + \lambda^{n-1} \cdot 0 = x_0$ is in the middle nucleus, for, by (5)

$$(\lambda^i x_i) \circ [x_0(\lambda^j x_j)] = (\lambda^i x_i) \circ [\lambda^j x_j(x_0 S^j)] = \lambda^i \lambda^j x_j(x_i S^j)(x_0 S^j)$$

while

$$[(\lambda^i x_i) \circ x_0] \circ \lambda^j x_j) = (\lambda^i x_i x_0) \circ (\lambda^j x_j) = \lambda^i \lambda^j x_j(x_i x_0) S^j.$$

We now propose to show that for the cases in which we will be interested, the elements named are the only members of the middle nucleus. For let

$$a = a_0 + \lambda a_1 + \cdots + \lambda^{n-1} a_{n-1} \tag{13}$$

be an element of the right nucleus, and let a_k be the last non-zero coefficient.

Then, in $(\lambda^k x)\lambda^{n-k}(a)$, the only term in λ^k will be the one arising from $\lambda^k a_k$. Using the distributive law, then, we must have

$$(\lambda^k x)\circ\delta a_k = [\delta(xS^{n-k})]\circ\lambda^k a_k = \lambda^k(xS^n)(\delta S^k)a_k \tag{14}$$

for all x in K. *This cannot occur unless* $S^n = I$ *and* δ *lies in some proper subfield of* K.

3. Division algebras. We now proceed to a discussion of those algebras $\Omega_{\delta,n,S}$ which are division algebras. Observe that $\Omega_{\delta,n,S}$ is a division algebra if and only if L_z is not a null form; i. e., if and only if there is no $z = z_0 + \lambda z_1 + \cdots + \lambda^{n-1}z_{n-1}$ such that $|L_z| = 0$. The only case to be considered here is the case $n = m$. We first prove the following

LEMMA. *When* $n = m$, *the determinant* $|L_z|$ *is a polynomial of degree at most* $n - 1$ *in* δ *with coefficients in the field* F, *for any* $z = z_0 + \cdots + \lambda^{n-1}z_{n-1}$, *where the* z_i*'s are in* K.

Proof. The determinant

$$|L_z| = \begin{vmatrix} z_0 & z_1 & \cdots & z_{n-1} \\ \delta(z_{n-1}S) & (z_0S) & \cdots & (z_{n-2}S) \\ \vdots & \vdots & & \vdots \\ \delta(z_1S^{n-1}) & \delta(z_2S^{n-1}) & \cdots & (z_0S^{n-1}) \end{vmatrix}$$

is clearly a polynomial of degree $n - 1$ in δ with coefficients in K. To show that the functions of the z_i's which give the coefficients are actually in F *for any set of* z_i*'s in* K, it is sufficient to show that

$$|L_{zS}| = |L_z|$$

where $|L_{zS}|$ is the determinant of the left multiplication for

$$z_0S + \lambda(z_1S) + \cdots + \lambda^{n-1}(z_{n-1}S),$$

for $|L_{zS}|$ is precisely the $n - 1$ degree polynomial $|L_z|$ in which each coefficient has been replaced by its image under S. But clearly, if $S^n = I$,

$$L_z = \begin{pmatrix} 0 & 0\cdots 0 & \delta^{-1} \\ 1 & 0\cdots 0 & 0 \\ 0 & 1\cdots 0 & 0 \\ \vdots & \vdots \quad \vdots & \vdots \\ 0 & 0\cdots 1 & 0 \end{pmatrix} \quad L_{zS} \quad \begin{pmatrix} 0 & 1\cdots 0 \\ 0 & 0\cdots 0 \\ \vdots & \vdots \quad \vdots \\ 0 & 0\cdots 1 \\ \delta & 0\cdots 0 \end{pmatrix} \tag{15}$$

and so $|L_z| = |L_{zS}|$.

We can now state and prove

THEOREM 2. *If $n = m$, then $\Omega_{\delta,n,S}$ is a division algebra if δ satisfies no polynomial of degree $n-1$ over F. For $n = m$ there always exist division algebras.*

Proof. By the lemma, we see that $|L_z|$ is a polynomial of degree $n-1$ over F, and hence, if δ satisfies no such polynomial then then $|L_z| = 0$ implies $z_0 = z_1 = \cdots = z_{n-1} = 0$, and hence Ω is a division algebra as claimed. For any n, since $[K:F] = n$, there are always elements satisfying no polynomial of degree less than n over F, and hence there always exist division algebras for any n.

Note that if n is prime, Ω is a division algebra if and only if $\delta \neq \delta S$, while if n is composite, Ω a division algebra implies $\delta \neq \delta S$.

4. Autotopism groups. We now come to the study of the autotopism groups of the algebras $\Omega_{\delta,n,S}$. The autotopism groups of those algebras which have been shown to be division algebras are of geometric interest, of course. The autotopisms of the other algebras are not hard to find given the results obtained for the division algebras, so they will be obtained as objects of algebraic interest. If any should be shown to be division algebras (unlikely!) then some more new planes and their collineation groups will have already been computed.

If Q, P, U are three non-singular linear transformations of Ω over $GF(p)$, then $\gamma(Q,P,U)$ will be said to be an *autotopism* of Ω if

$$PL_{zQ} = L_z U \tag{16}$$

for all z in Ω. The multiplication of two autotopisms is defined by

$$\gamma(Q,P,U)\cdot\gamma(Q_1,P_1,U_1) = \gamma(QQ_1,PP_1,UU_1). \tag{17}$$

We begin our examination of the autotopism group with the following

LEMMA. *Let Ω be any algebra with a unit. Then a is an element of the middle nucleus of Ω if and only if for every z in Ω*

$$L_a L_z = L_y \tag{18}$$

for some y in Ω.

Proof. The element a is in the middle nucleus of Ω if and only if $L_aL_z = L_{z\circ a}$ for all z in Ω. But if $L_aL_z = L_y$, then $1(L_aL_z) = aL_z = z\circ a = 1L_y = y$.

If we set $z = 1$ in (16), we get $PL_{1Q} = U$, so that we can write

$$PL_{zQ} = L_zPL_{1Q}. \tag{19}$$

This can be further rewritten

$$PL_z = L_{zQ^{-1}}PL_{1Q}. \tag{20}$$

Now, in (20), let $z = 1Q\circ a$ for a in the middle nucleus of Ω. Then we get

$$PL_{1Q\circ a} = L_{(1Q\circ a)Q^{-1}}PL_{1Q} = PL_aL_{1Q} \tag{21}$$

or

$$PL_a = L_{(1Q\circ a)Q^{-1}}P.$$

For any x in Ω, we can write

$$PL_aL_x = PL_{x\circ a} = L_{(1Q\circ a)Q^{-1}}PL_x = L_{(x\circ a)^{-1}}PL_{1Q} = L_{(1Q\circ a)Q^{-1}}L_{xQ^{-1}}PL_{1} \tag{22}$$

This yields

$$L_{(x\circ a)Q^{-1}} = L_{(1Q\circ a)Q^{-1}}L_{xQ^{-1}}, \tag{23}$$

for all x in Ω, so by the lemma, we know that $(1Q\circ a)Q^{-1}$ is in the middle nucleus whenever a is in the middle nucleus. From (21), then, we know that for any a in the middle nucleus, there is an element c in the middle nucleus such that

$$PL_a = L_cP. \tag{24}$$

With this result, we can study the form of the transformation P in the case of interest to us, namely where $N_r = N_m = K$. *From this point on, we will always assume that we are working with an algebra $\Omega_{\delta,n,s}$ satisfying this condition.* Write

$$P = (H_{ij}) \qquad i,j = 1,2,\cdots,n \tag{25}$$

where every H_{ij} is a linear transformation of K over $GF(p)$; i.e.,

$$H_{ij} = R(a_0) + TR(a_1) + T^2R(a_2) + \cdots + T^kR(a_k);\ a_i \text{ in } K. \tag{26}$$

Equation (24), then, can be written

$$(27)\qquad \begin{bmatrix} H_{11} & \cdots & H_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ H_{n1} & \cdots & H_{nn} \end{bmatrix} \begin{bmatrix} a & & & & \\ & (aS) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & (aS^{n-1}) \end{bmatrix} = \begin{bmatrix} c & & & & \\ & (cS) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & (cS^{n-1}) \end{bmatrix} \begin{bmatrix} H_{11} & \cdots & H_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ H_{n1} & \cdots & H_{nn} \end{bmatrix}.$$

But this is equivalent to the relations

$$(28)\qquad H_{ij}R(aS^{j-1}) = R(cS^{i-1})H_{ij} \text{ for } i, j = 1, 2, \cdots, n.$$

We quote here the following Lemma (Albert [1], Lemma 2):

Let H be a linear transformation over $GF(p)$ of K such that $HR(a) = R(c)H$ for every c of K where a is in K. Then

$$H = VR(h),$$

where h is in K and V is an automorphism over $GF(p)$ of K.

Applying the lemma, we see that (28) implies $H_{ij} = U_{ij}R(h_{ij})$ where U_{ij} is an automorphism of K and h_{ij} an element of K. Setting $H_{11} = VR(h_{11})$, we immediately see that $U_{ij} = VS^{j-i}$, or

$$(29)\quad P = \begin{bmatrix} VR(h_{11}) & SVR(h_{12}) & S^2VR(h_{13}) & \cdots & S^{n-1}VR(h_{1n}) \\ S^{-1}VR(h_{21}) & VR(h_{22}) & SVR(h_{23}) & \cdots & S^{n-2}VR(h_{2n}) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ S^{-n+1}VR(h_{n1}) & S^{-n+2}VR(h_{n2}) & S^{-n+3}VR(h_{n3}) & \cdots & VR(h_{nn}) \end{bmatrix}.$$

Now, in (19), let $z = \lambda$, let $1Q = u_0 + \lambda u_1 + \cdots + \lambda^{n-1}u_{n-1}$ and let $\lambda Q = a_0 + \lambda a_1 + \cdots + \lambda^{n-1}a_{n-1}$, and write

$$
\begin{pmatrix} H_{11} & \cdots & H_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ H_{n1} & \cdots & H_{nn} \end{pmatrix}
\begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ \delta(a_{n-1}S) & (a_1S) & \cdots & (a_{n-2}S) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \delta(a_1S^{n-1}) & \delta(a_2S^{n-1}) & \cdots & (a_0S^{n-1}) \end{pmatrix}
\tag{30}
$$

$$
= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ \delta & 0 & 0 & \cdots & 0 \end{pmatrix}
\begin{pmatrix} H_{11} & \cdots & H_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ H_{n1} & \cdots & H_{nn} \end{pmatrix}
\begin{pmatrix} u_0 & \cdots & u_{n-1} \\ \delta(u_{n-1}S) & \cdots & (u_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(u_1S^{n-1}) & \cdots & (u_0S^{n-1}) \end{pmatrix}.
$$

But this is equivalent to

$$
\begin{aligned}
&H_{i1}R(a_j) + H_{i2}R(a_{j-1}S) + \cdots + H_{ij}R(a_0S^j) \\
&\quad + H_{i,j+1}R(a_{n-1}S^{j+1})R(\delta) + \cdots + H_{i,n}R(a_{j-1}S^{n-1})R(\delta) = h_{i+1,1}R(u_j) \\
&\quad + H_{i+1,2}R(u_{j-1}S) + \cdots + H_{i+1,n}R(u_{j-1}S^{n-1})R(\delta) \\
&\text{for } i \neq n, \text{ and} \\
&H_{n1}R(a_j) + \cdots + H_{nn}R(a_{j-1}S^{n-1})R(\delta) = R(\delta)H_{11}R(u_j) \\
&\qquad + \cdots + R(\delta)H_{1n}R(u_{j-1}S^{n-1})R(\delta).
\end{aligned}
\tag{31}
$$

Notice that if it were known that *exactly* one element of the first row of P was non-zero, or some $h_{1k} \neq 0$, then, for $l \neq k$, letting j vary from 1 to n, we would have

$$H_{1l}R(a_{j-l+1}S^{l-1}) = H_{2,l+1(\mathrm{mod}\, n)}R(u_{j-l+2}S^{j-l+2}) = 0$$

or

$$H_{1l}R(a_{j-l+1}S^{l-1})\delta = H_{2,l+1(\mathrm{mod}\, n)}R(u_{j-l+2}S^{j-l+2})\delta = 0$$

so $h_{2,l+1(\mathrm{mod}\, n)} = 0$.

Continuing, we see that *exactly one element in each row or column is non-zero*, and the set of all non-zero elements lies on a "transversal"; i.e.,

$$h_{1i}, h_{2,i+1}, \cdots, h_{n,i-1} \neq 0, \text{ all other } h\text{'s} = 0. \tag{32}$$

To prove that exactly one element of the first row of P is non-zero, we need the following lemma which is a special case of a theorem in a paper of Albert. Lemma (Albert [1], p. 213, Theorem 4):

Let Ω be an algebra and $\gamma(Q, P, U)$ an autotopism of Ω. Then the mappings

$$x \to (x \circ 1P)P^{-1} \qquad (33)$$
$$x \to (1P \circ x)P^{-1}$$

induce automorphisms of the middle and right nuclei respectively.

Recall now that the algebras under consideration here all have their middle and right nuclei equal (isomorphic) to K. This implies that the two sets of all elements of the form $\{x \circ 1P\}$ and $\{1P \circ x\}$ must be the same, where x is in K. But if $z_0 + \lambda z + \cdots + \lambda^{n-1} z_{n-1} = z$, is an element of Ω such that the sets $\{x \circ z\}$ and $\{z \circ x\}$ are equal where x is in K, and z_i, z_j are both non-zero $(i < j)$ take an element y in K such that $yS^i \neq yS^j$. Then

$$y \circ z = \cdots \lambda^i z_i (yS^i) + \cdots + \lambda^j z_j (yS^j) + \cdots,$$

while

$$z \circ x = \cdots \lambda^i z_i x + \cdots + \lambda^j z_j x + \cdots,$$

for any x in K. So that when the j terms are equal, the i terms cannot be. But this cannot happen if z is to commute with the set $\{x\}$, x in K, so only one term of z can be non-zero. But if $z = 1P$, this implies that there is exactly one non-zero term in the first row of P as was asserted.

Let $G_{\delta,n,S}$ be the autotopism group of $\Omega_{\delta,n,S}$, and let $H_{\delta,n,S}$ be the subgroup of autotopisms containing P's of the form

$$P = \begin{bmatrix} VR(h_1) & & & & \\ & VR(h_2) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & VR(h_n) \end{bmatrix}. \qquad (34)$$

As a final result in this section we show that H is a *normal* subgroup of G. For let $\gamma(Q, P, U)$ be in H. Then, for any $\gamma(Q_1, P_1, U_1)$,

$$\begin{aligned} \gamma(Q_1^{-1}, P_1^{-1}, U_1^{-1})\gamma(Q, P, U)\gamma(Q_1, P_1, U_1) & \\ = \gamma(Q_1^{-1}QQ_1, P_1^{-1}PP_1, U_1^{-1}UU_1). & \end{aligned} \qquad (35)$$

But we have just seen that P_1 must be of the form indicated in (32), so P_1^{-1} must be of the form

$$h_{1,n-i}, h_{2,n-i+1}, \cdots, h_{n,n-i-1} \neq 0, \text{ all the other } h\text{'s} = 0. \qquad (36)$$

Clearly, then

$$P_1^{-1}PP_1 = \begin{pmatrix} VR(k_1) & & & \\ & VR(k_2) & & \\ & & \ddots & \\ & & & VR(k_n) \end{pmatrix}$$

for some set of k's. We state this result as a

LEMMA. *Let $G_{\delta,n,S}$ be the autotopism group of $\Omega_{\delta,n,S}$, and let $H_{\delta,n,S}$ be the subgroup of autotopisms with P as in* (34), *then H is normal in G.*

5. $\boldsymbol{\delta \neq \delta S}$, $\boldsymbol{n = m = 2}$. The first type of algebra we shall explicitly study is $\Omega_{\delta,n,S}$ where $\delta \neq \delta S$ and $n = m = 2$.

By Theorem 2, $\Omega_{\delta,2,S}$ is a division algebra ($\delta \neq S$). The equation to be investigated is (16). We will here be interested in the normal subgroup $H_{\delta,2,S} = H$ discussed at the end of Section 4. We can therefore write

$$\begin{aligned} &\begin{pmatrix} VR(h_0) & 0 \\ 0 & VR(h_1) \end{pmatrix} \begin{pmatrix} a_0 & a_1 \\ \delta(a_1S) & (a_0S) \end{pmatrix} \\ &= \begin{pmatrix} x_0 & x_1 \\ \delta(x_1S) & (x_0S) \end{pmatrix} \begin{pmatrix} VR(h_0) & 0 \\ 0 & VR(h_1) \end{pmatrix} \begin{pmatrix} u_0 & u_1 \\ \delta(u_1S) & (u_0S) \end{pmatrix} \end{aligned} \tag{37}$$

where $(x_0 + \lambda x_1)Q = a_0 + \lambda a_1$, and $1Q = u_0 + \lambda u_1$. Observe that, since $\gamma(I, R_a, R_a)$ is an autotopism for all a in K, where $R_a = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, we can examine the subgroup H_1 of H where $h_0 = 1$, and we do so in what follows. Multiplying we see that (37) is equivalent to the four equations:

$$\begin{aligned} a_0 &= (x_0V)u_0 + (x_1V)h_1\delta(u_1S) \\ a_1 &= (x_0V)u_1 + (x_1V)h_1(u_0S) \\ h_1\delta(a_1S) &= (\delta V)(x_1VS)u_0 + (x_0VS)h_1\delta(u_1S) \\ h_1(a_0S) &= (\delta V)(x_1VS)u_1 + (x_0VS)h_1(u_0S). \end{aligned} \tag{38}$$

Substituting the values obtained for a_0 and a_1 from the first two equations of (38) into the second two equations, we obtain (remembering that $S^2 = I$):

$$\begin{aligned} \delta h_1(h_1S)u_0 &= (\delta V)u_0 \\ (\delta S)h_1(h_1S)u_1 &= (\delta V)u_1. \end{aligned} \tag{39}$$

Clearly, if both u_0 and u_1 are non-zero, we quickly obtain the relation $\delta = \delta S$, which contradicts our hypothesis, so we know that, for $\gamma(Q, P, U)$ in H_1, $1Q = u_0$ or λu_1 since $u_0 \cdot u_1 = 0$. But $PL_{1Q} = U$, and it follows that the subgroup H_2 of H_1, containing autotopisms where

$$U = \begin{pmatrix} V_1R(a) & 0 \\ 0 & V_1R(b) \end{pmatrix}$$

is a normal subgroup of index one or two in H_1. This normal subgroup is completely determined by finding all solutions to the first equation of (39),

$$\frac{\delta V}{\delta} = h_1(h_1 S). \tag{40}$$

Things can be made yet simpler by considering the normal subgroup H_3 of H_2, in which $V = \mathrm{I}$. Then we need solve only

$$h_1(h_1 S) = 1. \tag{41}$$

But if $S^2 = 1$, as assumed, $h_1(h_1 S)$ is the norm of h_1 (written $N_S(h_1)$), and is automatically in F (also, every element of F is a norm). If ω is an element in K of order $q^2 - 1$ (i.e., a generator of the cyclic multiplicative group of K) then we need all solutions to

$$\omega^{\mu(1+q)} = 1 \text{ or } \mu(1+q) \equiv 0 \pmod{q^2 - 1}. \tag{41'}$$

There are $1 + q$ solutions to (41′), namely $\mu = 0, 1 - q, 2(1-q), \cdots, q(1-q)$.

Let A be the autotopism

$$\gamma\left(\begin{pmatrix} 1 & 0 \\ 0 & \omega^{1-q} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \omega^{1-q} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \omega^{1-q} \end{pmatrix}\right),$$

and B, the autotopism $\gamma(L_\omega, I, L_\omega)$. Then the order of A, $(o(A))$, is equal to $1 + q$ and $o(B) = q^2 - 1$. Together, they generate H_3, and further, by direct computation, every element in H_3 is uniquely expressible in the form $A^i B^j$, so $o(H_3) = (q^2 - 1)(q + 1)$. Now, there is a smallest power of T, say T^{k_0}, such that $(\delta T^{k_0})/\delta$ is in F. Since H_3 is normal in H_2 and H_3/H_2 is isomorphic to a cyclic subgroup of the group generated by T, H_2 is generated by H_3 and an autotopism

$$C = \gamma\left(T^{k_0}\begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix}, T^{k_0}\begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix}, T^{k_0}\begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix}\right).$$

Now C^{2r/k_0} will be in H_3, and so $o(H_2) = (q + 1)(q^2 - 1)2r/k_0$. Also H_2 was of index one or two in H_1, and

$$D = \left(\begin{pmatrix} 0 & S \\ SR(\delta) & 0 \end{pmatrix}, \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}, \begin{pmatrix} 0 & S \\ SR(\delta) & 0 \end{pmatrix} \right),$$

is in H_1 but not in H_2, therefore A, B, C, and D generate H_1 whose order is $(q+1)(q^2-1)4r/k_0$. Further computation shows that H is generated by H_1 and $E = \gamma(I, R_\omega, R_\omega)$, while G is generated, finally by H and

$$F = \left(\begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}, L_\lambda, L_\lambda \right).$$

We can therefore finally state

THEOREM 3. *When $n = m = 2$, $G_{\delta,2,S}$ is generated by the six elements A-F listed above, and its order is $(q+1)(q^2-1)^2 8r/k_0$. Also, G is solvable.*

Proof. Everything but the solvability of G has already been explicitly shown, and solvability follows from the examination of the sub-normal series defined in the determination of the group. Solvability in the case $n = 2$ has also been shown in [3].

6. $\mathfrak{S} \neq \mathfrak{S}S$, $n > 2$. In this case, as above, we wish to reduce from $G_{\delta,n,S}$ to $H_{\delta,n,S}$ and define H_1 as the subgroup of H where P is of the form

$$P = \begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & VR(h_{n-1}) \end{bmatrix}. \tag{42}$$

We can define H_1 since, as previously, $\gamma(I, R_a, R_a)$ is an autotopism for all a in K where

$$R_a = \begin{bmatrix} a & & & & \\ & a & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & a \end{bmatrix}.$$

If $a_0 + \lambda a_1 + \cdots + \lambda^{n-1} a_{n-1} = (x_0 + \lambda x_1 + \cdots + \lambda^{n-1} x_{n-1}) Q$ and $u_0 + \lambda u_1 + \cdots + \lambda^{n-1} u_{n-1} = 1Q$, we can write (16) as:

$$\begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & VR(h_{n-1}) \end{bmatrix} \begin{bmatrix} a_0 & \cdots & a_{n-1} \\ \delta(a_{n-1}S) & \cdots & (a_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(a_1S^{n-1}) & \cdots & (a_0S^{n-1}) \end{bmatrix}$$

$$= \begin{bmatrix} x_0 & \cdots & x_{n-1} \\ \delta(x_{n-1}S) & \cdots & (x_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(x_1S^{n-1}) & \cdots & (x_0S^{n-1}) \end{bmatrix} \begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & VR(h_{nn}) \end{bmatrix} \begin{bmatrix} u_0 & \cdots & u_{n-1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(u_1S^{n-1}) & \cdots & (u_0S^{n-1}) \end{bmatrix}. \tag{43}$$

From the first row we see

$$\begin{aligned} a_i = {} & (x_0V)u_i + h_1(x_1V)(u_{i-1}S) + \cdots + h_i(x_iV)(u_0S^i) \\ & + \delta h_{i+1}(x_{i+1}V)(u_{n-1}S^{i+1}) + \cdots + \delta h_{n-1}(x_{n-1}V)(u_{i+1}S^{n-1}), \end{aligned} \tag{44}$$

for all i, which proves that $Q = U = PL_{1Q}$. Equating the two entries in the second row, first column position, and using (44), we write

$$\begin{aligned} h_1\delta(a_{n-1}S) = {} & (\delta V)(x_{n-1}SV)u_0 + (x_0SV)\delta(u_{n-1}S)h_1 + \cdots \\ & + (x_{n-2}SV)(u_1S^{n-1})h_{n-1}\delta \\ = {} & h_1\delta[(x_0VS)(u_{n-1}S) + (h_1S)(x_1VS)(u_{n-2}S^2) + \cdots \\ & + (x_{n-1}VS)(u_0S^n)(h_{n-1}S)]. \end{aligned} \tag{45}$$

Equating coefficients of the different x_i's ((45) must hold for every element in Ω) we obtain:

$$\begin{aligned} h_2(u_{n-2}S^2) &= h_1(h_1S)(u_{n-2}S^2) \\ h_3(u_{n-3}S^3) &= h_1(h_2S)(u_{n-3}S^3) \\ &\cdot \\ &\cdot \\ &\cdot \\ h_k(u_{n-k}S^k) &= h_1(h_{k-1}S)(u_{n-k}S^k) \\ &\cdot \\ &\cdot \\ &\cdot \\ (\delta V)u_0 &= \delta h_1(h_{n-1}S)(u_0S^n). \end{aligned} \tag{46}$$

From the $(2,2)$ position we write

$$\begin{aligned} h_1(a_0S) = \delta V(x_{n-1}SV)u_1 + (x_0SV)h_1(u_0S) + (x_1SV)h_2\delta(u_{n-1}S^2) \\ + \cdots + (x_{n-2}SV)\delta h_{n-1}(u_2S^{n-1}) = h_1\{(x_0VS)u_0S) \\ + (x_1VS)[(\delta h_1)S](u_{n-1}S^2) + \cdots + (x_{n-1}VS)[(\delta h_{n-1})S](u_1S^n)\}, \end{aligned}$$

or

$$\begin{aligned} h_2\delta(u_{n-1}S^2) &= (\delta S)h_1(h_1S)(u_{n-1}S^2) \\ h_3\delta(u_{n-2}S^3) &= (\delta S)h_1(h_2S)(u_{n-2}S^3) \\ &\cdot \\ &\cdot \\ &\cdot \\ (\delta V)u_1 &= (\delta S)h_1(h_{n-1}S)(u_1S^n). \end{aligned} \tag{47}$$

In general, in the $(2,i)$ position for $i > 2$, we obtain from equating the coefficients of (x_1SV),

$$h_2u_{i-3}S^2 = h_1(h_1S)(u_{i-3}S^2). \tag{48}$$

Using (48) and the first equation of (46) we now have proven the following

LEMMA. *If any* $u_i \neq 0$, $i \neq n-1$, *then* $h_2 = h_1(h_1S)$.

Assume now that $u_{n-1} \neq 0$. The $(2, n-1)$ position yields (after equating the coefficients of $(x_{n-2}S)$),

$$h_1(h_{n-2}S)(\delta S) = h_{n-1}\delta \tag{49}$$

or, transforming (49) by S,

$$(h_1S)(h_{n-2}S^2)(\delta S^2) = (h_{n-1}S)(\delta S). \tag{49'}$$

The coefficient of $(x_{n-1}SV)$ in the $(2,n)$ position gives

$$h_1(h_{n-1}S)(\delta S)(u_{n-1}S^n) = (\delta V)u_{n-1}, \tag{50}$$

while the coefficient of $(x_{n-2}S^2V)$ in the $(3,n)$ position gives

$$h_2(h_{n-2}S^2)(\delta S^2)(u_{n-1}S^n) = (\delta V)u_{n-1}. \tag{51}$$

But (49′) and (50) imply $h_1(h_1S)(h_{n-2}S^2)(u_{n-1}S^n)(\delta S^2) = (\delta V)u_{n-1}$ which, together with (51) gives $h_2 = h_1(h_1S)$. But this last equation, together with the first equation of (47) gives $\delta = \delta S$, which contradicts our hypothesis in this section. *We know, then, that* $u_{n-1} = 0$.

We now turn to the coefficients of (x_1S^2V) in the $(3,i)$ position, and find, for all $j \neq n-1$, that we get

$$h_3(u_jS^3) = (h_1S^2)h_2(u_jS^3), \tag{52}$$

which, together with the previous lemma proves that if any $u_j \neq 0$ for $j \neq 1$, we can write $h_3 = h_1(h_1S)(h_1S^2)$. Continuing, we look at the coefficient of $(x_1S^{k-1}V)$ in the (k,i) position, $k \leqq n-1$, and obtain

$$h_k = h_{k-1}(h_1S^{k-1}). \tag{53}$$

But under the induction assumption, namely that $h_{k-1} = h_1(h_1S)\cdot\cdot\cdot(h_1S^{k-2})$, (53) implies that

$$h_k = h_1(h_1S)\cdot\cdot\cdot(h_1S^{k-1}). \tag{54}$$

Now, (54), together with (47) implies that if $u_j \neq 0$, for any $j = 2, 3, \cdot\cdot\cdot, n-2$, then $\delta = \delta S$, which is a contradiction. To date, then, we know that $u_j = 0$, when $j > 1$.

As a final computation, we propose to show that $u_1 = 0$. The coefficient of $(x_{n-2}S^2V)$ in the $(3,2)$ position gives

$$(\delta V)u_1 = h_2\delta(h_{n-2}S^2)(u_1S^n), \tag{55}$$

but, by (54), we can rewrite (55) as

$$(\delta V)u_1 = \delta(u_1S^n)h_1(h_1S)\cdot\cdot\cdot(h_1S^{n-1}).$$

Rewriting the last equation of (47) as

$$(\delta V)u_1 = (\delta S)h_1(h_1S)\cdot\cdot\cdot(h_1S^{n-1})(u_1S^n),$$

and examining the two equations, we obtain $\delta = \delta S$, a contradiction, unless $u_1 = 0$. We sum up all the consequences so far derived as follows:

THEOREM 4. *Let $\Omega_{\delta,n,S}$ be an algebra where $\delta \neq \delta S$ and if $n = m$, δ lies in no proper subfield of K, let $\gamma(Q,P,U)$ be an autotopism, and let $1Q = u_0 + \lambda u_1 + \cdot\cdot\cdot + \lambda^{n-1}u_{n-1}$, then $u_1 = \cdot\cdot\cdot = u_{n-1} = 0$; also, if*

$$P = \begin{bmatrix} V & & & & 0 \\ & VR(h_1) & & & \\ 0 & & \cdot & & \\ & & & \cdot & \\ & & & & VR(h_{n-1}) \end{bmatrix},$$

then $h_k = h_1(h_1S)\cdot\cdot\cdot(h_1S^{k-1})$.

At this point, we introduce the following notation; if $\Omega_{\delta,n,S}$ is the algebra under consideration, let

$$N_S(x) = x(xS)\cdot\cdot\cdot(xS^{n-1}) \tag{56}$$

for any x in K. The function $N_S(x)$, is a norm-like function, and if $n = m$, $N_S(x)$ will be exactly the norm of x with respect to the subfield F. This function will appear often in what follows.

Using Theorem 4, we can rewrite equation (16) as follows:

$$\begin{bmatrix} V & & & & \\ & VR(h) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & VR(b) \end{bmatrix} L_{a_0+\cdots+\lambda^{n-1}a_{n-1}} \tag{57}$$

$$= L_{x_0+\cdots+\lambda^{n-1}x_{n-1}} \begin{bmatrix} VR(u) & & & & \\ & VR(h)R(uS) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & VR(uS^{n-1})R(b) \end{bmatrix}$$

where $b = h(hS)\cdots(hS^{n-2})$, and where $a_i = (x_iV)h(hS)\cdots(hS^{i-1})(uS^i)$ if $i > 1$, and $a_0 = (x_0V)u$, $a_1 = (x_1V)(uS)h$. Equating the (i,i) positions of both sides, we obtain

$$\begin{aligned} h(hS)\cdots(hS^{i-2})(a_0S^{i-1}) \\ = (x_0S^{i-1}V)(uS^{i-1})h(hS)\cdots(hS^{i-2}) \end{aligned} \tag{58}$$

which is an identity. If $i < j$, the (i,j) position yields

$$h(hS)\cdots(hS^{i-2})(a_{j-i}S^{i-1}) = (x_{j-i}VS^{i-1})h(hS)\cdots(hS^{j-2})(uS^{j-1}), \tag{59}$$

which is also an identity. Finally, if $i > j$, we obtain

$$\begin{aligned} h(hS)\cdots(hS^{i-2})(a_{n+j-i}S^{i-1})\delta \\ = (\delta V)(x_{n+j-i}VS^{i-1})\cdot h(hS)\cdots(hS^{j-2})(uS^{j-1}), \end{aligned} \tag{60}$$

or

$$\delta h(hS)\cdots(hS^{n+j-2})(uS^{n+j-1}) = (\delta V)h(hS)\cdots(hS^{j-2})(uS^{j-1}), \tag{60$'$}$$

or

$$\frac{\delta V}{\delta} = \left[\frac{(uS^n)}{u}N_S(h)\right] S^{j-1} \tag{60$''$}$$

This last equation must hold for every $j < n - 1$, and we restate it as two conditions:

a) $$\frac{\delta V}{\delta} = \frac{uS^n}{u} N_S(h),$$

b) $$\left[\frac{\delta V}{\delta}\right] S = \frac{\delta V}{\delta}.$$

At this point, we must distinguish between the cases $n = m$ and $m > n$. In the case $n = m$, $uS^n/u = 1$ for all u in K, and $[N_S(h)]S = N_S(h)$ for all h in K, so we must solve

$$\frac{\delta V}{\delta} = N_S(h) \tag{61}$$

for some h in K, or alternatively, $\delta V/\delta$ must be in F. Let H_2 be the normal subgroup of H_1, where $V = I$. Then H_2 has two generators,

$$A_1 = \gamma\left(\begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix}, \begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix},\right.$$
$$\left.\begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix}\right),$$

where $k = \omega^{q-1}$, ω primitive, so that k generates the elements of norm 1 in K, and

$$B_1 = \gamma(L_\omega, I, L_\omega).$$

Clearly, $A_1B_1 = B_1A_1$, and

$$o(A) = q^{n-1} + q^{n-2} + \cdots + q + 1 = \frac{q^n - 1}{q - 1},$$

so,

$$o(H_2) = \frac{q^n - 1}{q - 1}(q^n - 1).$$

Now, H_1 is generated by H_2 and an element

$$C_1 = \gamma\left(\begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & VR(c) \end{bmatrix}, \begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & VR(c) \end{bmatrix}, \begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & VR(c) \end{bmatrix}\right)$$

$(c = h_1(h_1S) \cdots (h_1S^{n-2}))$ where $V\colon x \to x^{p^{k_0}}$ is the least power of p with the property that $\delta^{p^{k_0}}\delta^{-1}$ is in F, and where

$$\frac{\delta^{p^{k_0}}}{\delta} = N_S(h_1).$$

We know, then, that

$$o(H_1) = \frac{nr}{k_0}\,\frac{(q^n-1)^2}{(q-1)}.$$

Also, $H_{\delta,n,S}$ is generated by H_1 and $D_1 = \gamma(I, R_\omega, R_\omega)$, as above, so

$$o(H) = \frac{nr}{k_0}\,\frac{(q^n-1)^3}{(q-1)}.$$

Finally, $G_{\delta,n,S}$ is generated by H and $E_1 = \gamma(L_\lambda, S^{-1}, L_\lambda)$ whose powers are seen to give all n cosets where

$$P = \begin{bmatrix} 0 & \cdots & 0 & VR(h_{1t}) & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & VR(h_{2,t+1}) & \cdots & 0 \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ 0 & \cdots & VR(h_{n,t-1}) & 0 & 0 & \cdots & 0 \end{bmatrix}$$

so that we can finally state the following

THEOREM 5. *Let $n = m > 2$, and choose δ so that it lies in no subfield of K. Then, $\Omega_{\delta,n,S}$ is a division algebra whose autotopism group is generated by the element A_1—E_1 above. The autotopism group is solvable and has order*

$$\frac{n^2 r}{k_0} \cdot \frac{(q^n-1)^3}{q-1}.$$

Proof. Every statement has already been shown except the solvability of G which follows from the construction of the group as in the case $n=m=2$.

The case where $m>n$ is somewhat different since, in (60″), we have neither $uS^n=u$, nor $N_S(h)$ in F, in general. In this case, let H_2 be the normal subgroup of H_1 where $V=I$. Then we need solve

$$\frac{(uS^n)}{u} \cdot N_S(h) = 1. \tag{61}$$

To begin with, let ω generate the cyclic multiplicative group of K. Let $u=\omega^{\mu}$. Then (60) asks for all solutions to the equation

$$N_S(h) = \omega^{(\mu-\mu q^n)}. \tag{62}$$

If $h=\omega^{\nu}$, (62) is equivalent to

$$\omega^{\nu}(\omega^{\nu q}) \cdots (\omega^{\nu(q^{n-1})}) = \omega^{\mu(1-q^n)},$$

or

$$\nu(1+q+\cdots+q^{n-1}) \equiv \mu(1-q^n) \bmod (q^m-1). \tag{63}$$

Let $\nu=\mu(1-q)$, and ν gives a solution. Therefore, for any u, there is an autotopism in H_2 such that $1Q=u+\lambda\cdot 0+\cdots+\lambda^{n-1}\cdot 0$. H_2, then, is generated by

$$A_2=\gamma\left(\begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \ddots & & \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix}, \begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \ddots & & \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix}, \begin{bmatrix} 1 & & & & \\ & k & & & \\ & & \ddots & & \\ & & & & k(kS)\cdots(kS^{n-2}) \end{bmatrix}\right)$$

where k generates the cyclic subgroup of the multiplicative group of elements x such that $N_S(x)=1$, and

$$B_2 = \gamma\left[\begin{bmatrix} 1 & & & & \\ & h & & & \\ & & \ddots & & \\ & & & h & \cdots\ (hS^{n-2}) \end{bmatrix} L_\omega,\ \begin{bmatrix} 1 & & & & \\ & h & & & \\ & & \ddots & & \\ & & & h & \cdots\ (hS^{n-2}) \end{bmatrix},\right.$$

$$\left.\begin{bmatrix} 1 & & & & \\ & h & & & \\ & & \ddots & & \\ & & & h & \cdots\ (hS^{n-2}) \end{bmatrix}\right] L_\omega$$

where $h = \omega/\omega^q$. The order of A_2 depends on the relationship between n and m, and must divide $q^m - 1$. Let $o(A_2) = t$. Then $o(H_2) = t(q^m - 1)$. Also, H_1 is generated by H_2 and

$$C_2 = \gamma\left[V \begin{bmatrix} 1 & & & & \\ & l & & & \\ & & \ddots & & \\ & & & l & \cdots\ (lS^{n-2}) \end{bmatrix} L_{\omega^\alpha},\right.$$

$$\left. V \begin{bmatrix} 1 & & & & \\ & l & & & \\ & & \ddots & & \\ & & & l & \cdots\ (lS^{n-2}) \end{bmatrix},\ V \begin{bmatrix} 1 & & & & \\ & l & & & \\ & & \ddots & & \\ & & & l & \cdots\ (lS^{n-2}) \end{bmatrix} L_{\omega^\alpha} \right]$$

where $V : x \to x^{p^{k_0}}$ is the smallest power of T such that $\delta V/\delta$ satisfies conditions (a) and (b) following (60″) and where (60″) is satisfied by letting $\omega^\alpha = u$ and $l = h$. The order of H_1 is $o(H_2) \cdot (mr/k_0) = (tmr/k_0)(q^m - 1)$. Finally $H_{\delta,n,s}$ is generated by H_1 and

$$D_2 = \gamma(I, R_\omega, R_\omega).$$

We will now prove that when $n \neq m$, $H_{\delta,n,s} = G_{\delta,n,s}$, and be finished with this case.

Lemma. *If $m > n$ and $\gamma(Q, P, U)$ is an autotopism of $\Omega_{\delta,n,s}$, then P is of the form*

$$\begin{bmatrix} VR(h_0) & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & VR(h_{n-1}) \end{bmatrix}.$$

Proof. In equations (31), if we interpret the H_{ij}'s as in (29), then for all $i \neq n$, we get $h_{in}(a_j S^{n-1}) = 0$ for all j. But since some $a_j \neq 0$, this shows that $h_{nn} \neq 0$. This completes the proof of the lemma.

We have, then,

THEOREM 6. *Let $n > 2$, $m \neq n$ and let δ be such that $\delta \neq \delta S$. Then $G_{\delta,n,S}$ is generated by A_2—D_2 above, has order $t(mr/k_0)(q^m - 1)^2$, and is solvable.*

7. $\delta = \delta S$. In order to be able to use Theorem 1 in this case, we must restrict ourselves to the condition $m \neq n (\delta = \delta S)$. Some of these algebras might be division algebras, but whether they are or not, their autotopism groups will be determined here as of general algebraic interest.

Since the right and middle nuclei are both of the form

$$x_0 + \lambda \cdot 0 + \cdots + \lambda^{n-1} \cdot 0,$$

all x_0 in K, our previous assertions about the nature of the transformation P still hold. Notice that the lemma preceding Theorem 6 was proved independently of any properties of δ, and is therefore true in the case under consideration here also, so $H = G$. If we define H_1 to be the subgroup $G_{\delta,n,S}$ satisfying (42), then H_1 is determined by the solution of equations (43) and (44). Let H_2 be the normal subgroup of H_1 in which $V = I$, and we will solve (43) where $V = I$.

The first step will be to show that

$$h_k = h(hS) \cdots (hS^{k-1})$$

is implied by *any* u_i being non-zero. To see this, observe that the coefficients of $(x_1 V S^{k-1})$ in the (k, i) position imply the same results for u_{n-1} as for all the other u_i's when $\delta = \delta S$.

In the (i, j) position where $i \leqq j$, (43) gives

$$\begin{aligned} h_{i-1}(a_{j-i} S^{i-1}) = \delta(x_{n-i+1} S^{i-1}) u_{j-1} & \\ + h_1 \delta (x_{n-i+2}) S^{i-1} (u_{j-2} S) & \\ + \cdots + h_{j-1} (x_{n-i+j} S^{i-1}) (u_0 S^{j-1}) & \\ + \delta h_j (x_{n-i+j+1} S^{i-1}) (u_{n-1} S^j) + \cdots & \end{aligned} \tag{64}$$

where the subscripts are all reduced modulo n. But

$$a_{j-i} = x_0 u_{j-i} + h_1 x_1 (u_{j-i-1} S) + \cdots \\ + h_{j-i} x_{j-i} (u_0 S^{j-i}) + \delta h_{j-i+1} x_{j-i+1} (u_{n-1} S^{j-i+1}) + \cdots$$

and since $\delta S = \delta$, the δ's exactly cancel when a_{j-i} is transformed by S^{i-1}, and multiplied by h_{i-1}, and when the coefficients of $x_k S^{i-1}$ are equated. For the coefficient of $(x_{n-i+k} S^{i-1})$, we obtain

$$h_{i-1}(h_{n+k-i} S^{i-1})(u_{j-k} S^{n+k-1}) = (u_{j-k} S^{k-1}) h_{k-1} \tag{65}$$

if $k - i < 0$, and

$$h_{i-1}(h_{k-i} S^{i-1})(u_{j-k} S^{k-1}) = (u_{j-k} S^{k-1}) h_{k-1}$$

if $k - 1 \geqq 0$. This latter is an indentity, since

$$h_{k-i} = h(hS) \cdots (hS^{k-i-1}), \text{ so } (h_{k-1} S^{i-1}) = (hS^{i-1}) \cdots (hS^{k-2}).$$

In the previous case, though, we obtain

$$\frac{u_{j-k} S^{k-1}}{u_{j-k} S^{n+k-1}} = (hS^{k-1})(hS^{k}) \cdots (hS^{n+k-2}),$$

which is the equation

$$\frac{u_{j-k}}{u_{j-k} S^{n}} = N_S(h) \tag{66}$$

operated on by S^{k-1}. If $i > j$, the δ's are also seen to cancel out, and the same equations are arrived at. What we have demonstrated, then, is the following

LEMMA. *Let* $\delta = \delta S$, $n \neq m$, *and* $V = I$. *Then equation* (43) *is satisfied by* $1Q = u_0 + \lambda u_1 + \cdots + \lambda^{n-1} u_{n-1}$ *and* $h_k = h(hS) \cdots hS^{k-1})$, $h_0 = 1$, *if and only if*

$$u_i = N_S(h)(u_i S^n) \tag{67}$$

for all i.

Now, if $u_i \neq 0$, $u_j \neq 0$, we have $u_i^{1-q^n} = u_j^{1-q^n}$, or $u_j = u_i \cdot \mu$ where $\mu^{1-q^n} = 1$. Also, if $u_i \neq 0$, $h = u_i^{1-q^n} \cdot \nu$ where $N_S(\nu) = 1$ gives all solutions for h. The number of solutions μ of $\mu^{1-q^n} = 1$ is $q^{(n,m)} - 1$. The number of solutions ν of $N_S(\nu) = 1$ depends on p, n, and m, and these solutions form a cyclic subgroup of order $t \mid (q^m - 1)$, and are generated by ν_0. Now, for any k, $1 \leqq k \leqq n$, there are $tC_k{}^n (q^m - 1)(q^{(m,n)} - 1)^{k-1}$ autotopisms in H_2 with *exactly* k non-zero u_i's, and hence, the order of H_2 is

$$t(q^m - 1) \sum_{k=1}^{n} (q^{(m,n)} - 1)^{k-1} C_k{}^n,$$

where $C_k{}^n$ is the number of ways of choosing k distinct objects from n, with no regard to order. The group H_1 is generated by H_2 and a solution of (43) using the smallest power of T, say T^{k_0}, used in any autotopism of G. Finally, $G_{\delta,n,S}$ is generated by H_1 and the autotopism $\gamma(I, R_\omega, R_\omega)$, so that we have proved the following

THEOREM 7. *In the case* $n \neq m$, $\delta = \delta S$, *the autotopism group,* $G_{\delta,n,S}$ *is of order*

$$(tmn/k_0)(q^m - 1)^2 \sum_{k=1}^{n} (q^{(m,n)} - 1)^{k-1} C_k{}^n.$$

8. Isomorphic and isotopic algebras. The next topic to be discussed is the quesion of isomorphism between two algebras. $\Omega_{\delta,n,S}$ will be said to be *isomorphic* to $\Omega_{\epsilon,n,S}$ if there is a non-singular linear map V of $\Omega_\delta \to \Omega_\epsilon$ such that $(x \circ y)V = xV * yV$, for any x, y in Ω_δ, where $\circ$ is the product operation in Ω_δ and $*$, in Ω_ϵ. We can prove the following.

THEOREM 8. *Suppose* $n \neq m$ *or if* $n = m$, *neither* δ *nor* ϵ *lies in a proper subfield of* K, *then* Ω_δ *is isomorphic with* Ω_ϵ *if and only if there is an automorphism* A, *of* K *and an element* h *of* K *such that*

$$(68) \qquad (\delta A) = h(hS) \cdots (hS^{n-1})\epsilon$$

and, if $n \neq 2$,

$$hS^n = h.$$

Proof. Observe first that if Ω_δ and Ω_ϵ are isomorphic, then by Theorem 1, the right and middle nuclei of both algebras are isomorphic, and are all elements of the form $x_0 + \lambda \cdot 0 + \cdots + \lambda^{n-1} \cdot 0$. Clearly, the isomorphism V must map the right nucleus of Ω_δ onto the right nucleus of Ω_ϵ and can be considered an automorphism, A, of K when acting on elements of the nucleus. Let $\Omega_{\epsilon,n,S}$ be generated by K and the element λ_0, where $\lambda_0{}^n = \epsilon$, and let

$$(69) \qquad \lambda V = a_0 + \lambda_0 a_1 + \cdots + \lambda_0{}^{n-1} a_{n-1}.$$

Then, for x in K, $(x\lambda)V = (xA) * (a_0 + \lambda_0 a_1 + \cdots + \lambda_0{}^{n-1} a_{n-1}) = (\lambda(xS))V = (\lambda V) * (xSA) = (a_0 + \lambda_0 a_1 + \cdots + \lambda_0{}^{n-1} a_{n-1}) * (xSA)$. But if $a_k \neq 0$ the $\lambda_0{}^k$ term gives $xSA = xS^kA$ for all x in K which is impossible unless $k = 1$, hence $V: \lambda \to \lambda_0 h$. Now, V is an automorphism, so we know that $\lambda^i V = (\lambda V)^i = (\lambda_0 h)^i = \lambda_0{}^i h(hS) \cdots (hS^{i-1})$, and

$$\lambda^n V = \lambda_0{}^n h(hS) \cdots (hS^{n-1}) = \epsilon h(hS) \cdots (hS^{n-1}) = \delta A.$$

Unless $n = 2$,

$$\begin{aligned}(\lambda^{n-1}\circ\lambda^{2})V&=[\lambda_0^{n-1}h(hS)\cdots(hS^{n-2})]*[\lambda_0^{2}h(hS)]\\&=\lambda_0\epsilon h(hS)(hS)^{2}\cdots(hS^{n})=(\lambda\delta)V=\lambda_0 h(\delta A)\\&=\lambda_0 h\epsilon h(hS)\cdots(hS^{n-1})\end{aligned}$$

which implies $h=hS^n$. Notice that when $m=n$, $S^n=I$ and this last condition is automatically satisfied. On the other hand if A is an automorphism of K such that $\delta A=\epsilon h(hS)\cdots(hS^{n-1})$ and $hS^n=h$ (if $n>2$), then define $V\colon x\rightarrow xA$ for x in K, and $V\colon \lambda\rightarrow\lambda_0 h$, $\lambda^i\rightarrow\lambda_0^{i}h(hS)\cdots(hS^{i-1})$, $V\colon \lambda^i x\rightarrow(\lambda^i V)(xA)$. To show that V is an isomorphism, we need only verify that $(\lambda^i x_i)V*(\lambda^j x_j)V=[(\lambda^i x_i)\circ(\lambda^j x_j)]V$. But $(\lambda^i x_i)V=\lambda^i V(x_iA)$ $=\lambda_0^{i}h(hS)\cdots(hS^{i-1})(x_iA)$, so

$$(\lambda^i x_i)V*(\lambda^j x_j)V=\lambda_0^{i+j}h(hS)(hS^2)\cdots(hS^{i+j-1})(x_iS^jA)(x_jA)$$

if $i+j<n$, in which case

$$[(\lambda^i x_i)\circ(\lambda^j x_j)]V=\lambda_0^{i+j}(x_iS^jA)(x_jA)h(hS)\cdots(hS^{i+j-1}).$$

If $i+j>n$, we have

$$\begin{aligned}&\lambda_0^{i+j-n}h(hS)(hS^2)\cdots(hS^{i+j-1})(x_iAS^j)(x_jA)\epsilon\\&\quad=\lambda_0^{i+j-n}h(hS)\cdots(hS^{i+j-n-1})(x_iAS^j)(x_jA)\epsilon h(hS)\cdots(hS^{n-1}).\end{aligned}$$

This equality holds if and only if

$$h(hS)\cdots(hS^{i+j-n-1})=(hS^n)(hS^{n+1})\cdots(hS^{i+j-1}),$$

which is true if $hS^n=h$ as assumed. If $i+j=n$, we get the condition $\delta A=\epsilon h(hS)\cdots(\mathrm{h}S^{n-1})$ which was assumed. The case $n=2$ has been discussed by Albert [1] and we do not include the proof here.

Finally, we come to the question of isotopy beween the division algebras of Theorem 2.

Two algebras $\Omega_{\delta,n,S}$ and $\Omega_{\epsilon,n,S}$ are said to be *isotopic* if there are three non-singular linear mappings of $\Omega_{\epsilon,n,S}$ onto $\Omega_{\delta,n,S}$, Q, P, U, such that

$$PL_{xQ}=L'_xU, \tag{70}$$

where L_x is left multiplication in $\Omega_{\delta,n,S}$, and L'_x, in $\Omega_{\epsilon,n,S}$. The geometrical interest in isotopy between algebras is due to the well-known theorem (Albert [2], Theorem 6) asserting that *two division algebras coordinatize isomorphic projective planes if and only if the algebras are isotopic.* The investigations in this section will allow us to prove:

THEOREM 9. *Let $n=m$, and let δ and ϵ be two elements of K, neither of which lies in any proper subfield of K. Then $\Omega_{\delta,n,S}$ and $\Omega_{\epsilon,n,S}$ are isotopic if and only if they are isomorphic.*

Notice that the argument concerning the form of P given in Section 4 is applicable in this case also, and, multplying the isotopism by an appropriate *autotopism of* $\Omega_{\delta,n,S}$, we can assume that P is of the form

$$\begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & VR(h_{n-1}) \end{bmatrix}.$$

We can rewrite (70), then as

$$\begin{bmatrix} V & & & & \\ & VR(h_1) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & VR(h_{n-1}) \end{bmatrix} \begin{bmatrix} a_0 & \cdots & a_{n-1} \\ \delta(a_{n-1}S) & \cdots & (a_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(a_1S^{n-1}) & \cdots & (a_0S^{n-1}) \end{bmatrix}$$

(70′)

$$\begin{bmatrix} x_0 & \cdots & x_{n-1} \\ \epsilon(x_{n-1}S) & \cdots & (x_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \epsilon(x_1S^{n-1}) & \cdots & (x_0S^{n-1}) \end{bmatrix} P \begin{bmatrix} u_0 & \cdots & u_{n-1} \\ \delta(u_{n-1}S) & \cdots & (u_{n-2}S) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \delta(u_1S^{n-1}) & \cdots & (u_0S^{n-1}) \end{bmatrix}$$

where $1Q = u_0 + \lambda u_1 + \cdots + \lambda^{n-1}u_{n-1}$. To begin with, if $n = 2$, then (70′) can be written

$$\begin{pmatrix} V & 0 \\ 0 & VR(h) \end{pmatrix} \begin{pmatrix} a_0 & a_1 \\ \delta(a_1S) & (a_0S) \end{pmatrix}$$

$$= \begin{pmatrix} x_0 & x_1 \\ \epsilon(x_1S) & (x_0S) \end{pmatrix} \begin{pmatrix} V & 0 \\ 0 & VR(h) \end{pmatrix} \begin{pmatrix} u_0 & u_1 \\ \delta(u_1S) & (u_0S) \end{pmatrix}$$

which is equivalent to the equations

$$\begin{aligned} a_0 &= (x_0V)u_0 + (x_1V)h\delta(u_1S) \\ a_1 &= (x_0V)u_1 + (x_1V)h(u_0S) \\ h\delta(a_1S) &= (\epsilon V)(x_1SV)u_0 + (x_0SV)h\delta(u_1S) \\ h(a_0S) &= (\epsilon V)(x_1SV)u_1 + (x_0SV)h(u_0S). \end{aligned}$$

These yield the equations (remember $S^2 = I$):

$$\begin{aligned} (\epsilon V)u_0 &= h(hS)\delta u_0 \\ (\epsilon V)u_1 &= h(hS)(\delta S)u_1. \end{aligned} \tag{71}$$

Since either u_0 or u_1 must be non-zero, (71) implies the conditions of Theorem 8, and hence the truth of the theorem for $n = 2$.

If $n > 2$, the same argument used above (i. e., using the $(2, n-1)$, $(2, n)$, and $(3, n)$ positions) shows that $u_{n-1} = 0$. But looking at the coefficients of $(x_1 S^k V)$ in the $(k+1, i)$ positions shows, also as above, that

$$h_k = h(hS) \cdots (hS^{k-1})$$

holds.

To prove Theorem 9, if any $u_i \neq 0$ for $i \neq 0$, or $n-1$, the (n, n) position of (70′) yields

$$\begin{aligned} h_{n-1}(a_0 S^{n-1}) = (\epsilon V)(x_2 V S^{n-1}) h_1(u_{n-2} S) + \cdots \\ + (\epsilon V)(x_{n-1} V S^{n-1}) h_{n-2}(u_1 S^{n-1}) + (x_0 V S^{n-1}) h_{n-1}(u_0 S^{n-1}). \end{aligned}$$

But equating the coefficients of x_i for $i \neq 0$, 1, we obtain the equality (since $S^n = I$)

$$h_{n-1}(h_i S^{n-1})(\delta S^{n-1}) = (\epsilon V) h_{i-1}.$$

But $h_{n-1}(h_i S^{n-1}) = N_S(h) \cdot h_{i-1}$, so we have $(\epsilon V) = N_S(h)(\delta S^{n-1})$, and transforming by S, we get $(\epsilon V S) = N_S(h)\delta$. If $u_0 \neq 0$, the $(n-1)$ position gives, $h_{n-1}\delta(a_1 S^{n-1}) = (\epsilon V) u_0 (x_1 S^{n-1} V) + \cdots$, and equating the coefficients of $(x_1 S^{n-1} V)$, we obtain $h_{n-1}\delta(h S^{n-1}) = (\epsilon V)$, or

$$(\epsilon V) = \delta N_S(h).$$

In either case, the conditions of Theorem 8 are satisfied, completing the proof of Theorem 9.

UNIVERSITY OF CHICAGO,
INSTITUTE FOR DEFENSE ANALYSES.

REFERENCES.

[1] A. A. Albert, "On the collineation groups of certain non-Desarguesian planes," *Portugaliae Mathematica*, vol. 18 (1959), pp. 207-224.

[2] ———, "Finite division algebras and finite planes," *Proceedings of Symposia in Applied Mathematics*, vol. X (1960), pp. 53-70.

[3] D. R. Hughes, "Collineation groups of non-Desarguesian planes II," *American Journal of Mathematics*, vol. 82 (1960), pp. 113-119.

[4] D. R. Hughes and E. Kleinfeld, "Seminuclear extensions of Galois fields," *American Journal of Mathematics*, vol. 82 (1960), pp. 389-392.

[5] G. Pickert, *Projective Ebenen*, Berlin, 1955.

THE FIRST AND SECOND FUNDAMENTAL THEOREMS OF LIE FOR LIE PSEUDO GROUPS.*

By A. M. Rodrigues.

Introduction. In this paper we prove two theorems (Theorems 3.1 and 3.2) which establish a one to one correspondence between regular Lie pseudo groups acting on M (see Definition 2.4) and certain sheaves of germs of local vector fields of M. This correspondence seems to play for Lie pseudo groups a similar role to the correspondence between Lie groups and Lie algebras. The local form of these theorems were in essence already known to S. Lie [6]. Making use of the theory of jets of C. Ehresmann the author proves these theorems in a global form and under hypotheses which appear to be the most natural ones; in particular, no assumption of analyticity is made throughout this paper. At the same time some natural questions concerning the relationship between the pseudo group and vector fields defined by one parameter local groups belonging to the pseudo group are clarified. For the definition and basic facts about jets we refer the reader to [2].

In Section 4 it is shown that, under very mild hypotheses, the prolongation of a Lie pseudo group is again a Lie pseudo group. This is used in Section 5 to derive the equations of structure of E. Cartan [1] for some prolongation of the pseudo group. In this paragraph we follow closely the approach of M. Kuranishi ([4], Chapter IV).

The writer had his attention called to the infinitesimal pseudo group (see Definition 3.1) attached to a Lie pseudo group, by a recent paper of P. Libermann [5]. Some of the ideas of that paper are used here freely.

The author wishes to express his grateful thanks to Professors M. Kuranishi and D. C. Spencer for many valuable discussions on the subject of this paper.

1. Extensions of a vector field.

Let M be a connected differentiable manifold of class C^∞ [1] and of dimension n. A local diffeomorphism f of M is a diffeomorphism from an open set of M onto an open set of M; we shall denote by $U(f)$ and $V(f)$ respectively the domain and the range of f. Let

---

* Received October 27, 1961.

[1] All differentiable manifolds and differentiable maps considered in this paper are assumed to be of class C^∞.

$j^s(M)$ be the manifold of all s-jets of local diffeomorphisms of M and denote by $\alpha: j^s(M) \to M$ and $\beta: j^s(M) \to M$ the projections which map each jet $X \in j^s(M)$ respectively into its source $\alpha(X)$ and target $\beta(X)$.

The s-jet of a local diffeomorphism at the point $x \in U(f)$ will be denoted by $j_x{}^s f$ or simply $j^s f$ whenever the source of the jet is clearly determined by the context. Given two coordinate neighborhoods U and V in M with coordinates $(x^1, \cdots, x^n)$ and $(y^1, \cdots, y^n)$ we define a system of coordinates in $(\alpha \times \beta)^{-1}(U \times V) \subset j^s(M)$ in the following way: For any s-jet

$$X \in (\alpha \times \beta)^{-1}(U \times V) \text{ let } f = (\cdots, y^i(x^1, \cdots, x^n), \cdots)$$

$i = 1, \cdots, n$, be a local diffeomorphism of M such that $X = j_{x_0}{}^s f$ then, the coordinates of X are $(x_0{}^i, y^i(x_0{}^i, \cdots, x_0{}^n), p_{j_1, \cdots j_\nu})$, $1 \leqq \nu \leqq s$, $1 \leqq j_k \leqq n$, where $p_{j_1, \cdots j_\nu}$ is the partial derivative $\dfrac{\partial y^i}{\partial x^{j_1} \cdots \partial x^{j_\nu}}$ computed at the point $(x_0{}^1, \cdots, x_0{}^n)$ (see [2]).

By a local vector field Θ defined in the open set $U \subset M$ we shall understand a local cross section of M into its tangent bundle $T(M)$, defined over U; we shall use the functional notation $\Theta(x)$ to denote the value of the vector field Θ at the point $x \in U$. The s-jet of source x of the cross section $\Theta: U \to T(M)$ will be denoted by $j_x{}^s\Theta$. $j_x{}^s\Theta$ is an element of the vector space $T_x{}^s$ of all s-jets at x of local vector fields defined in a neighborhood of $x \in M$. Let f_t be the one parameter local group of local transformations defined by Θ; then we define the extension $f_t{}^s$ of f_t by the formula $f_t{}^s(X) = j^s f_t \circ X$; $f_t{}^s$ is clearly a one parameter local group acting on $\beta^{-1}(U)$.

Definition 1.1. The s-th prolongation Θ^s of Θ is the vector field defined in $\beta^{-1}(U)$ by the one parameter local group $f_t{}^s$.

Given $x \in M$ let $A(x)$ be the submanifold of $j^s(M)$ of all jets whose source is x; $A(x) = \alpha^{-1}(x)$. From Definition 1.1 it is clear that the vector $\Theta^s(X)$ is tangent to $A(\alpha(X))$ for every $X \in \beta^{-1}(U)$.

For each $Z \in j^s(M)$ let $R(Z): A(\alpha(Z)) \to A(\beta(Z))$ be the map defined by $R(Z)(X) = X \circ Z^{-1}$. $R(Z)$ is a diffeomorphism of $A(\alpha(Z))$ onto $A(\beta(Z))$.

PROPOSITION 1.1. *Let* $X \in \beta^{-1}(U)$ *and* $Z \in j^s(M)$ *be two jets such that* $\alpha(X) = \alpha(Z)$ *then,*

$$R(Z)_*(\Theta^s(X)) = \Theta^s(R(Z)(X)),$$

where $R(Z)_*$ *indicates the map induced by* $R(Z)$ *on tangent vectors.*

Proof. $\Theta^s(X)$ and $\Theta^s(R(Z)(X))$ are respectively the tangent vectors at $t = 0$ to the parametrized curves $j^s f_t \circ X$ and $j^s f_t \circ (X \circ Z^{-1})$. The proposition follows from the equation $R(Z)(j^s f_t \circ X) = j^s f_t \circ (X \circ Z^{-1})$.

Let $I^s(x)$, or simply $I(x)$ when there is no possibility of mistake, denote the s-jet at $x \in M$ of the identity map of M.

PROPOSITION 1.2. *If* Θ_1 *and* Θ_2 *are two local vector fields defined in the open set* $U \subset M$ *such that* $j_x{}^s\Theta_1 = j_x{}^s\Theta_2$ *then* $\Theta_1{}^s(I(x)) = \Theta_2{}^s(I(x))$.

Proof. Let $(x^1, \cdots, x^n)$ be a system of coördinates defined in a neighborhood $V \subset U$ of x and consider the system of coordinates

$$(x^1, \cdots, x^n, y^1, \cdots, y^n, p_{j_1}, \cdots {}_{j_\nu})$$

defined in a neighborhood of $I(x)$ in $j^s(M)$.

Let $\Theta_1 = \sum_{i=1}^{n} \theta^i(x^1, \cdots, x^n) \dfrac{\partial}{\partial x^i}$ be the local expression of Θ in the local coordinates $(x^1, \cdots, x^n)$ and let $f_t = (\cdots, f^i(x^1, \cdots, x^n, t), \cdots)$ be the one parameter local group generated by Θ_1. We have

$$f^i(x,t) = x^i + \theta^i(x,t)t + O(t^2)$$

$$\frac{\partial f^i(x,t)}{\partial x^j} = \delta_j{}^i + \frac{\partial \theta^i(x)}{\partial x^j} t + O(t^2)$$

$$\frac{\partial f^i(x,t)}{\partial x^{j_1}, \cdots, \partial x^{j_\nu}} = \frac{\partial \theta^i(x)}{\partial x^{j_1}, \cdots, \partial x^{j_\nu}} t + O(t^2).$$

A simple computation shows that the local expression of $\Theta_1{}^s$ at the point $I(x)$ in the coordinate system (x, y, p) is:

$$(1.1) \qquad \Theta_1{}^s(I(x)) = \sum_{i=1}^{n} \theta^i(x) \frac{\partial}{\partial y^i} + \sum \frac{\partial \theta^i(x)}{\partial x^{j_1}, \cdots, \partial x^{j_\nu}} \frac{\partial}{\partial p_{j_1}, \cdots {}_{j_\nu}}.$$

Proposition 1.2 follows immediately from Formula (1.1).

PROPOSITION 1.3. *If* Θ_1 *and* Θ_2 *are 2 local vector fields defined in the open set* $U \subset M$ *and* $j_x{}^s\Theta_1 = j_x{}^s\Theta_2$ *for a point* $x \in U$ *then* $\Theta_1{}^s(X) = \Theta_2{}^s(X)$ *for any jet* $X \in j^s(M)$ *such that* $\beta(X) = x$.

Proof. By Proposition 1.1 and 1.2 we have,

$$\begin{aligned}\Theta_1{}^s(X) = \Theta_1{}^s(R(X^{-1})I(\beta(X))) &= R(X^{-1})_*\Theta_1{}^s(I(\beta(X))) \\ &= R(X^{-1})_*\Theta_2{}^s(I(\beta(X))) = \Theta_2{}^s(X).\end{aligned}$$

For any $X \in j^s(M)$ let T_X be the tangent space of the manifold $A(\alpha(X))$ at the point X. It follows from Proposition 1.3 that for any $X \in j^s(M)$ there exists a natural map $\lambda^s(X) : T^s{}_{\beta(X)} \to T_X$. Given a vector $\tau \in T^s{}_{\beta(X)}$ and a local vector field Θ such that $j_x{}^s\Theta = \tau$ then

$$\lambda^s(X)(\tau) = \Theta^s(X).$$

PROPOSITION 1.4. *$\lambda^s(R(Z)(X)) = R(Z)_* \circ \lambda^s(X)$ for any $X, Z \in j^s(M)$ with $\alpha(X) = \alpha(Z)$.*

Proof. Proposition 1.4 follows from the definition of the map $\lambda^s(X)$ and from Proposition 1.1.

PROPOSITION 1.5. *$\lambda^s(X)$ is a linear isomorphism of $T^s{}_{\beta(X)}$ onto T_X.*

Proof. By Proposition 1.4 it suffices to consider the case $X = I(x)$ for any $x \in M$. Keeping the notation as in Proposition 1.2, the components of the vector $j_x{}^s\Theta_1$ with relation to the natural basis defined in $T_x{}^s$ by the system of coordinates $(x^1, \cdots, x^n)$ are $\theta^i(x)$, $\dfrac{\theta^i(x)}{\partial_x{}^{j_1}, \cdots, \partial_x{}^{j_\nu}}$, $1 \leqq i \leqq n$, $1 \leqq \nu \leqq s$, $1 \leqq j_k \leqq n$. This together with Formula (1.1) shows that $\lambda^s(I(x))$ is an injective linear map. To show that $\lambda^s(I(x))$ maps $T_x{}^s$ onto $T_{I(x)}$, let $\tau \in T_{I(x)}$ be a vector with components $\tau^i, \tau^i{}_{j_1, \cdots, j_\nu}$ with relation to the basis $\dfrac{\partial}{\partial y_i}, \dfrac{\partial}{\partial p^i{}_{j_1, \cdots, j_\nu}}$ of $I(x)$, and let $\theta^i(x)\, i = 1, \cdots, n$, be n functions defined in a neighborhood of $x \in M$ such that the value of $\theta^i(x)$ at the point x is τ^i and moreover $\dfrac{\theta^i(x)}{\partial_x{}^{j_1}, \cdots, \partial_x{}^{j_\nu}} = \tau^i{}_{j_1, \cdots, j_\nu}$, the partial derivatives being computed at x. If Θ_1 is the vector field $\sum_{i=1}^n \theta^i(x) \dfrac{\partial}{\partial_x{}^i}$ then, by formula (1), $\Theta_1{}^s(I(x)) = \tau$.

We shall often denote $\lambda^s(I(x))$ simply by $\lambda(x)$.

PROPOSITION 1.6. *If Θ_1 and Θ_2 are two vector fields defined in the open set $U \subset M$ then,*

$$[\Theta_1, \Theta_2]^s = [\Theta_1{}^s, \Theta_2{}^s].$$

Proof. Let f_t and g_u be respectively the one parameter local group generated by Θ_1 and Θ_2. Since

$$[\Theta_1, \Theta_2] = \lim_{t \to 0} \frac{1}{t} (\Theta_2 - f_{t*}(\Theta_2)), \text{ [7, p. 8]},$$

we have for any $X \in \beta^{-1}(U)$,

$$\begin{aligned} [\Theta_1, \Theta_2]^s(X) &= \lambda(X)\,(j^s{}_{\beta(X)}[\Theta_1, \Theta_2]) \\ &= \lambda(X)\,(j^s{}_{\beta(X)}(\lim_{t \to 0} \frac{1}{t}(\Theta_2 - f_{t*}\Theta_2))) \\ &= \lambda(X)\,(\lim_{t \to 0} \frac{1}{t}\,(j^s{}_{\beta(X)}\Theta_2 - j^s{}_{\beta(X)}(f_{t*}\Theta_2))) \end{aligned}$$

$$= \lim_{t\to 0} \frac{1}{t}(\lambda(X)(j^s{}_{\beta(X)}\Theta_2) - \lambda(X)(j^s{}_{\beta(X)}(f_{t*}\Theta_2)))$$
$$= \lim_{t\to 0} \frac{1}{t}(\Theta_2{}^s(X) - (f_{t*}\Theta_2)^s(X)).$$

But for a fixed t, $f_{t*}\Theta_2$ is the vector field generated by the one parameter group (the parameter is u) $f_t g_u f_t^{-1}$ [7, p. 7]; then, $(f_{t*}\Theta_2)^s$ is the vector field generated by $(f_t \circ g_u \circ f_t^{-1})^s = f_t^s \circ g_u^s \circ f_t^{s^{-1}}$. Therefore $(f_{t*}\Theta_2)^s = (f_t^s)_*\Theta_2^s$ and

$$[\Theta_1, \Theta_2]^s(X) = \lim_{t\to 0}\frac{1}{t}[\Theta_2{}^s(X) - (f_t{}^s)_*\Theta_2{}^s(X)] = [\Theta_1{}^s, \Theta_2{}^s](X).$$

2. Lie pseudo groups. Given two local homeomorphisms of f, g of M, we say that $f \circ g$ is defined if $V(g) \subset U(f)$.

Definition 2.1. A pseudo group operating in M is a set Γ of local diffeomorphisms of M such that:

1) If $f, g \in \Gamma$ and $f \circ g$ is defined then $f \circ g \in \Gamma$
2) If $f \in \Gamma$ then $f^{-1} \in \Gamma$
3) The identity map of M is in Γ.

Let $j^s(\Gamma)$ be the set of all s-jets of elements of Γ.

Definition 2.2. Γ is a Lie pseudo group acting on M if there exists an integer $s \geqq 0$, called the order of Γ, such that:

1) There exists a structure of differentiable manifold in $j^s(\Gamma)$ which makes $j^s(\Gamma)$ a submanifold of $j^s(M)$.
2) A local diffeomorphism f of M is in Γ if and only if $j_x{}^s f \in j^s(\Gamma)$ for all $x \in U(f)$.

Consider a family U_i of open sets of M and put $U = \bigcup U_i$; from Definition 2.2 it is clear that a local diffeomorphism f defined in U is in the Lie pseudo group Γ if and only if the restriction of f to each U_i is in Γ.

Let Γ be a Lie pseudo group acting on M. For $x \in M$ put $A'(x) = A(x) \cap j^s(\Gamma)$. We shall use the same notation for the map α and for its restriction to $j^s(\Gamma)$.

PROPOSITION 2.1. *$A'(x)$ is a submanifold regularly imbedded in $j^s(\Gamma)$ and the dimension of $A'(x)$ is independent of $x \in M$.*

Proof. It suffices to show that the restriction of the map α to $j^s(\Gamma)$ is of maximal rank at every point of $j^s(\Gamma)$. Take $X \in j^s(\Gamma)$ and put $x_0 = \alpha(X)$;

let $f \in \Gamma$ be such that $j_{x_0}^s f = X$. The map $\sigma: U(f) \rightarrow j^s(\Gamma)$ defined by $\sigma(x) = j_x^s f$ is a cross section with respect to α and $\sigma(x_0) = X$. This proves that α is of maximal rank at X.

Since the maps α and β are of maximal rank at every point of $j^s(\Gamma)$ and since the map $(X, Y) \rightarrow X \circ Y$ is differentiable, as can be readily proved, $j^s(\Gamma)$ is a differentiable groupoid in the sense of [3].

Definition 2.3. A vector field Θ defined in an open set $U \subset M$ is a Γ vector field if the transformations f_t of the one parameter local group generated by Θ belong to Γ for every t.

PROPOSITION 2.2. *A necessary and sufficient condition for Θ to be a Γ vector field is that the restriction of Θ^s to $j^s(\Gamma)$ is tangent to $j^s(\Gamma)$.*

Proof. From the definition of Θ^s it is clear that the condition is necessary. Conversely, if $X \in j^s(\Gamma)$ and the parametrized curve $j^s f_t \circ X$ is tangent to $j^s(\Gamma)$ at each of its points then, $j^s{}_{\beta(X)} f_t \in j^s(\Gamma)$ for all t and therefore $f_t \in \Gamma$.

PROPOSITION 2.3. *If Θ_1 and Θ_2 are Γ vector fields defined in the open set $U \subset M$ then, for any constants α_1, α_2, $\alpha_1\Theta_1 + \alpha_2\Theta_2$ and $[\Theta_1, \Theta_2]$ are Γ vector fields.*

Proof. Proposition 2.3 follows immediately from Propositions 1.6 and 2.2.

We shall denote by $\Theta(\Gamma)$ the sheaf over M of germs of Γ-vector fields; by Proposition 2.3 $\Theta(\Gamma)$ is a sheaf of Lie algebras. Let $j_x^s\Theta(\Gamma)$ be the set of all s-jets of source x of Γ vector fields.

PROPOSITION 2.4. *A necessary and sufficient condition for a vector field Θ defined in the open set $U \subset M$ to be a Γ vector field is that $j_x^s\Theta \in j_x^s\Theta(\Gamma)$ for all $x \in U$.*

Proof. The condition is obviously necessary. Conversely if $j_x^s\Theta \in j_x^s\Theta(\Gamma)$ for all $x \in U$ then, by Propositions 1.2 and 2.2, $\Theta^s(I(X))$ is tangent to $j^s(\Gamma)$. By Proposition 1.1 Θ^s is tangent to $j^s(\Gamma)$ at all points where it is defined and therefore, by Proposition 2.2, Θ is a Γ-vector field.

By Proposition 2.2, for $X \in j^s(\Gamma)$, $\lambda^s(X)$ maps $j^s{}_{\beta(X)}\Theta(\Gamma)$ into the tangent space T'_X of the manifold $A'(\alpha(X))$ at the point X.

Definition 2.4. Γ is a regular Lie pseudo group of order s if $\lambda^s(X)$ maps $j^s{}_{\beta(X)}\Theta(\Gamma)$ onto T'_X for every $X \in j^s(\Gamma)$.

Clearly, for Γ to be a regular Lie pseudo group it suffices that $\lambda^s(I(x))$ maps $j_x^s\Theta(\Gamma)$ onto $T'_{I(x)}$ for every $x \in M$.

Given a Lie subgroup G of $GL(n,R)$, let Γ be the pseudo group of all local diffeomorphisms f of R^n whose Jacobian matrices at each point $x \in U(f)$ is in G. Γ is a Lie pseudo group of order 1 and it can be easily proved that Γ is regular.

We want now to show that the Lie pseudo groups considered by Kuranishi [4] are always regular, although this result shall not be used in the remaining of the paper. First, we give a slightly different definition of the prolongation of a vector field and show that it is equivalent to Definition 1.1.

Let ϕ_t be a one parameter family of diffeomorphisms of M, satisfying the following conditions: 1) The domain of each transformation ϕ_t contains the open set $U \subset M$. 2) ϕ_t is defined for t varying in an open interval containing the origin of the real line. 3) ϕ_t depends differentiably on t. 4) ϕ_0 is the identity map of U. The family ϕ_t generates a vector field Θ defined in U in the following way: For $x \in U$, $\Theta(x)$ is the tangent vector at x of the parametrized curve $t \to \phi_t(x)$. We now define the extension $\tilde{\Theta}^s$ of Θ as the vector field defined in $\beta^{-1}(U) \subset j^s(M)$ by putting, for each $X \in \beta^{-1}(U)$, $\tilde{\Theta}^s(X) =$ tangent vector at X of the parametrized curve $t \to j^s{}_{\beta(X)}\phi_t \circ X$. It is easy to verify that the extension $\tilde{\Theta}^s$ does not depend on the choice of the family ϕ_t that generates Θ and, in particular, if we use in place of ϕ_t the one parameter local group f_t generated by Θ we see that $\tilde{\Theta}^s = \Theta^s$.

PROPOSITION 2.5. *Let Γ be a Lie pseudo group. If ϕ_t belongs to Γ for every t then Θ is a Γ vector field.*

Proof. By the definition of $\tilde{\Theta}^s$, its restriction to $j^s(\Gamma)$ is tangent to $j^s(\Gamma)$; since $\tilde{\Theta}^s = \Theta^s$, by Proposition 2.2 Θ is a Γ vector field.

Now we assume that M is a real analytic manifold and that the elements of Γ are bi-analytic maps; we say in this case that Γ is an analytic Lie pseudo group. Let ρ be the projection into the first factor of the product $M \times M$ and let $j^s(M \times M, M, \rho)$ denote the set of all s-jets of local cross sections, with respect to the projection ρ, of M into $M \times M$. Denote by $k\colon j^s(\Gamma) \to j^s(M \times M, M, \rho)$ the map defined in the following way: Given $f \in \Gamma$, $k(j_{x_0}{}^s f)$ is the s-jet of source x_0 of the cross section $x \to (x, f(x))$. Clearly k is an injection map; put $F = k(j^s(\Gamma))$.

Definition 2.5. The analytic Lie pseudo group Γ is involutive if the partial differential equation $F \subset j^s(M \times M, M, \rho)$ is in involution at every integral point, in the sense of [4].

PROPOSITION 2. 6. *An involutive Lie pseudo group is regular.*

Proof. Take a vector $\tau \in T'_{I(x)}$ and let $\gamma(t)$ be an analytic parametrized curve contained in $A'(x)$ whose tangent vector at $I(x)$ is τ. By the theorem of the analytic dependence on the initial conditions of the solutions of a partial differential equation which is in involution there exists an analytic family ϕ_t of bi-analytic local maps of M into M such that: 1) $\phi_t \in \Gamma$ for every t; 2) The parametrized curves $\gamma(t)$ and $t \rightarrow j_x{}^s\phi_t$ are equal. 3) For every t, ϕ_t is defined in a set containing a fixed open neighborhood U of x and ϕ_0 is the identity map of U. Let Θ be the vector field generated by ϕ_t; by Proposition 2. 5 Θ is a Γ vector field and by construction $\lambda(x)(j_x{}^s\Theta) = \Theta^s(I(x)) - \bar{\Theta}^s(I(x)) = \tau$ therefore, Γ is regular.

3. Infinitesimal Lie pseudo groups.

Definition 3. 1. An infinitesimal Lie pseudo group E, acting on the differentiable manifold M, is a sheaf of germs of local vector fields of M satisfying the conditions:

1) If Θ_1 and Θ_2 are sections of E over the open set $U \subset M$ then, for any real numbers α_1, α_2, $\alpha_1\Theta_1 + \alpha_2\Theta_2$ and $[\Theta_1, \Theta_2]$ are also sections of E.

2) Denote by $j_x{}^sE$ the vector space of s-jets of source x of sections of E defined in a neighborhood of x. Condition 2) says that there exists an $s \geqq 0$ such that a vector field Θ defined in the open set $U \subset M$ is a section of E if and only if $j_x{}^s\Theta \in j_x{}^sE$ for every $x \in U$.

3) The dimension of the vector space $j_x{}^sE$ is independent of $x \in M$.

The integer s is called the order of E.

We proved in Section 2 that if Γ is a *regular* Lie pseudo group then the sheaf $\Theta(\Gamma)$ has properties 1), 2) and 3) of Definition 3. 1. Therefore, we have

THEOREM 3. 1. *The sheaf of germs of* Γ *vector fields of a 'regular' Lie pseudo group* Γ *is an infinitesimal Lie pseudo group.*

We want now to prove that conversely, given an infinitesimal Lie pseudo group E there exists a regular Lie pseudo group Γ such $\Theta(\Gamma) = E$. To construct Γ we define a distribution S in $j^s(M)$ in the following way: to each point $X \in j^s(M)$ we associate the vector space $S_X = \lambda(X)(j^s{}_{\beta(X)}E)$. By condition 3 of Definition 3. 1 the dimension of S_X is independent of $X \in j^s(M)$. By Proposition 1. 4, $R(Z)_*S_X = S_{R(Z)(X)}$ for every Z and X such that

$\alpha(Z) = \alpha(X)$. It is a simple matter to prove that S is a distribution of class C^∞; from Proposition 1.6 it follows that S is involutive.

Let $\mathcal{J}_x$ be the maximal integral manifold of S which contains $I(x)$ and put $G = \bigcup_{x \in M} \mathcal{J}_x$.

PROPOSITION 3.1. *If $X, Y \in G$, and $X \circ Y^{-1}$ is defined, then $X \circ Y^{-1} \in G$.*

Proof. The map $R(Y)$ sends integral manifolds of S into integral manifolds of S. Since $Y \in \mathcal{J}_{\alpha(Y)}$ and $R(Y)(Y) = I(\beta(Y))$ we have $R(Y)(\mathcal{J}_{\alpha(Y)}) = \mathcal{J}_{\beta(Y)}$. Since $X \in \mathcal{J}_{\alpha(Y)}$, $R(Y)(X) = X \circ Y^{-1} \in \mathcal{J}_{\beta(Y)}$ and therefore $X \circ Y^{-1} \in G$.

As a corollary of Proposition 3.1 we have that if $X \in G$ then $X^{-1} \in G$.

We shall show that G has a natural structure of differentiable manifold such that with this structure G is a submanifold of $j^s(M)$. Let $I \subset j^s(M)$ be the set of all jets $I^s(x)$, $x \in M$; I is a submanifold of $j^s(M)$. Given $X \in G$ it is possible to find a coordinate neighborhood U of X in $j^s(M)$ coordinates $(y^1 \cdots y^r)$ and a coordinate neighborhood V of $I^s(\alpha(X))$ with coordinates $(x^1 \cdots x^r)$ such that:

1) For all values of the constants $a^1 \cdots a^t$ which are smaller, in absolute value, than a fixed number $\epsilon > 0$ the slices $y^1 = a^1 \cdots y^t = a^t$, $x^1 = a^1 \cdots x^t = a^t$ are integral manifolds of S contained in the same maximal integral manifold of S, (see [8], Corollary 1, page 12).

2) $I \cap V$ is the set defined by $x^{n+1} = \cdots = x^r = 0$.

3) $y^1(X) = \cdots = y^r(X) = 0$, $x^1(I^s(\alpha(X))) = \cdots = x^r(I^s(\alpha(X))) = 0$.
Let $W \subset U$ be the set of points defined by $y^{n+1} = \cdots = y^t = 0$. Clearly $W \subset G$. We take $(y^1 \cdots y^n, y^{t+1} \cdots y^r)$ as a system of coordinates in W. With these systems of coordinates G becomes a submanifold of $j^s(M)$ and the restriction of the map α to G is everywhere of maximum rank hence, G is a differentiable groupoid.

Let Γ be the pseudo group of all local diffeomorphisms f of M such that $j_x^s f \in G$ for all $x \in U(f)$.

PROPOSITION 3.2. *There exists a neighborhood U of the set $I \subset G$ such that for every $X \in U$ there exists $f \in \Gamma$, defined in a neighborhood of $\alpha(X)$, belonging to a one parameter local group of elements of Γ, such that $j^s_{\alpha(X)} f = X$.*

Proof. Since the map $X \to I(\alpha(X))$ is of maximum rank at each point of G it is clearly enough to show that for every $x \in M$ there exists a neighborhood of $I(x)$ in $\mathcal{J}_x$ with the required property. By definition of $\mathcal{J}_x$, given a tangent vector τ to $\mathcal{J}_x$ at the point $I(x)$ there exists a vector field Θ defined in a neighborhood V of x such that Θ^s is tangent to G at all points where it is defined and $\Theta^s(I(x)) = \tau$. Let f_t be the one parameter local group transformations defined by Θ. Then, $j_x{}^s f_t \in G$ for every t and $x \in V$, therefore $f_t \in \Gamma$. By a known theorem, varying τ in a neighborhood of zero in the tangent space of $\mathcal{J}_x$ at the point $I(x)$, the orbits of $I(x)$ under all one parameter groups so defined cover a neighborhood of $I(x)$ in $\mathcal{J}_x$. But the orbit of $I(x)$ under the one parameter local group $f_t{}^s$ (see Section 1) is the parametrized curve $t \to j_x{}^s f_t$; this proves Proposition 3.2.

PROPOSITION 3.3. *The differentiable groupoid G is generated by any neighborhood U of the set I in the sense that given $X \in G$ there exists a sequence $X_1 \cdots X_k$ of elements of U such that $X = X_1 \circ X_2 \cdots \circ X_k$.*

Proof. Let H be the subgroupoid of G generated by U. Since G is connected it suffices to show that H is open and closed in G. To show that H is open it is enough to prove that each set $B(x) = H \cap \mathcal{J}_x$ is open in $\mathcal{J}_x$. Take $X \in B(x)$ and put $V = \mathcal{J}_{\beta(X)} \cap U$; V is an open set in $\mathcal{J}_{\beta(X)}$. Clearly, $V \circ X$ is an open neighborhood of X in $\mathcal{J}_x$ and $V \circ X \subset B(x)$. To show that H is closed we introduce the equivalence relation in $G: X_1 \sim X_2$ if $X_1 \circ X_2 \in H$. Each equivalence class is entirely contained in some $\mathcal{J}_x$ and the equivalence classes define a partition of each $\mathcal{J}_x$ in open sets; hence, each equivalence class is also a closed set in $\mathcal{J}_x$. Since $B(x)$ is the equivalence class of $I(x)$, $B(x)$ is closed in $\mathcal{J}_x$; therefore, H is closed in G.

THEOREM 3.2: *Γ is a regular Lie pseudo group and E is the sheaf of germs of Γ vector fields.*

Proof. By Propositions 3.2 and 3.3 given $X \in G$ there exists $f \in \Gamma$ such that $j^s{}_{\alpha(X)} = f$; this and Proposition 3.1 show that Γ is a Lie pseudo group. Let now f_t be a one parameter group of local transformations belonging to Γ and let Θ be the vector field generated by f_t. Since $\Theta^s(I(x))$ is tangent to G at $I(x)$ it follows that $j_x{}^s\Theta$ belongs to $j_x{}^s E$ for every x where Θ is defined; hence, Θ is a section of E. Conversely, if Θ is a section of E then by construction of G, Θ^s restricted to G is tangent to G which proves that Θ is a Γ vector field.

We shall now answer the following question: How are two regular Lie pseudo groups which have the same infinitesimal Lie pseudo group related?

PROPOSITION 3.4. *If Γ is a regular Lie pseudo group then the connected component $\tilde{j}^s(\Gamma)$ of the set I in $j^s(\Gamma)$ is a differentiable groupoid.*

Proof. Consider the sheaf $\Theta(\Gamma)$ of Γ vector fields and let Γ_0 be the Lie pseudo group constructed as in Section 2, with the sheaf $\Theta(\Gamma)$. We shall prove that $j^s(\Gamma_0) = \tilde{j}^s(\Gamma)$. By Proposition 3.3 there exists a connected neighborhood U of I in $\tilde{j}^s(\Gamma)$ such that each point in U is the jet of a local diffeomorphism belonging to the one parameter local group of some Γ vector field; clearly $U \subset \tilde{j}^s(\Gamma)$. Since $\tilde{j}^s(\Gamma)$ is generated by U we have $j^s(\Gamma_0) \subset \tilde{j}^s(\Gamma)$. $\tilde{j}^s(\Gamma)$ is an open sub-manifold of $j^s\Gamma)$, and the manifolds $j^s(\Gamma_0)$ and $\tilde{j}^s(\Gamma)$ have the same dimension; therefore, $j^s(\Gamma_0)$ is open in $\tilde{j}^s(\Gamma)$. To finish the proof it suffices to show that $j^s(\Gamma_0)$ is also closed in $\tilde{j}^s(\Gamma)$. To prove that $j^s(\Gamma_0)$ is closed in $\tilde{j}^s(\Gamma)$ it is enough to show that for every $x \in M$, $A'_0(x) = A(x) \cap j^s(\Gamma_0)$ is closed in $A'(x) = A(x) \cap \tilde{j}^s(\Gamma)$. Take $Z \in A'(x)$, $Z \notin A'_0(x)$ and let $\mathcal{J}_Z$ be the maximal integral manifold of S (in the notation of Section 2) which contains Z. We claim that $\mathcal{J}_Z \subset j^s(\Gamma)$; in fact, $R(Z)_*\mathcal{J}_Z = \mathcal{J}_{I(\beta(Z))} \subset j^s(\Gamma)$ and Z itself is in $j^s(\Gamma)$. But $\mathcal{J}_Z$ is connected and $Z \in \tilde{j}^s(\Gamma)$ hence $\mathcal{J}_Z \subset j^s(\Gamma)$. $\mathcal{J}_Z$ is then an open neighborhood of Z in $A'(x)$ which does not meet $A'_0(x)$, for two distinct maximal integral manifolds of an involutive distribution have no common point. Therefore, $A'_0(x)$ is closed in $A'(x)$.

Definition 3.2. The regular Lie pseudo group Γ_0 is called the connected component of the identity, of order s, of the regular Lie pseudo group Γ. Γ is connected of order s if $j^s(\Gamma)$ is connected.

The proof of Proposition 3.4, also gives a proof of

PROPOSITION 3.5. *Two regular Lie pseudo groups of the same order acting on M which have the same infinitesimal Lie pseudo group have the same connected component of the identity of order s.*

We have also proved the following propositions:

PROPOSITION 3.6. *If Γ is a regular Lie pseudo group of order s then there exists a neighborhood U of the set I in $j^s(\Gamma)$ such that every jet $X \in U$ is the jet of an element $f \in \Gamma$ belonging to a one parameter local group of Γ.*

PROPOSITION 3.7. *If Γ is a regular and connected Lie pseudo group of order s then $j^s(\Gamma)$ is generated, as a differentiable groupoid, by any neighborhood of the set $I \subset j^s(\Gamma)$.*

4. Prolongation of a Lie pseudo group. Let Γ be a Lie pseudo group of order s acting on a manifold M. For each $f \in \Gamma$ let $p^s f$ be the local diffeomorphism of $j^s(\Gamma)$ defined as follows: The domain of definition of $p^s f$ is

$\beta^{-1}(U(f))$ and for $X \in \beta^{-1}(U(f))$, $p^s f(X) = j^s f \circ X$. Let $p^s\Gamma$ be the set of all local diffeomorphisms ϕ of $j^s(\Gamma)$ such that for each $X \in U(\phi)$ there exists a neighborhood $V \subset U(\phi)$ of X and an element $f \in \Gamma$ with $p^s f | V = \phi | V$. Clearly p^s is a pseudo group acting on $j^s(\Gamma)$.

Definition 4.1. $p^s\Gamma$ is the s-th prolongation of Γ.

We want to show that, imposing mild conditions on the Lie pseudo group Γ, $p^s\Gamma$ is also a Lie pseudo group. Let U be a coordinate neighborhood of M with coordinates $x^1 \cdots x^n$ and consider the differential forms ϕ^j, $\phi^j{}_{j_1\cdots j_\nu}$, $1 \leqq \nu \leqq s-1$, $1 \leqq j$, $j_1 \cdots j_\nu \leqq s$, defined in $j^s(U)$ (see [3],[2], Chapter IV, § 2).

If $X = (0, 0, p^j{}_{j_1\cdots j_\nu})$, $Y = (0, 0, q^j{}_{j_1\cdots j_\nu})$ are two s-jets whose source and target are the origin of R^n, put $X \circ Y = (0, 0, \phi^j{}_{j_1\cdots j_\nu}(p, q))$, $1 \leqq \nu \leqq s$; $\phi^j{}_{j_1\cdots j_\nu}(p, q)$ are polynomials in $p^j{}_{j_1\cdots j_k}$, $q^j{}_{j_1\cdots j_k}$ for $1 \leqq k \leqq \nu$, $1 \leqq j, j_1 \cdots j_\nu \leqq n$. Put $X^{-1} = (0, 0, \bar{p}_{j_1\cdots j_\nu}(p))$ and $\phi^j{}_{j_1\cdots j_\nu k}{}^{k_1\cdots k_\sigma}(p, q) = \dfrac{\partial \phi^j{}_{j_1\cdots j_\nu}(p, q)}{\partial q^k{}_{k_1\cdots k_\sigma}}$. The expression of the forms $\phi^j, \phi^j{}_{j_1\cdots j_\nu}$ in the system of coordinates $(x, y, p^j{}_{j_1\cdots j_\nu})$ of $j^s(U)$ are:

4.1 $$\begin{aligned} &\phi^j = \bar{p}_k{}^j(p)\, dy^k \\ &\phi^j{}_{j_1\cdots j_\nu} = \phi^j{}_{j_1\cdots j_\nu k}{}^{k_1\cdots k_\sigma}(p, p)\,(dp^k{}_{k_1\cdots k_\sigma} - p^k{}_{k_1\cdots k_\sigma k'}\, dy^{k'}). \end{aligned}$$

Let $\rho_s{}^{s+k}\colon j^{s+k}(U) \to j^s(U)$ be the map which sends each jet $X \in j^{s+k}(U)$ into its induced jet of order s. For $\nu \leqq s-1$ the forms $\phi^j, \phi^j{}_{j_1\cdots j_\nu}$ defined in $j^{s+k}(U)$ are the image under $(\rho_s{}^{s+k})^*$ of the corresponding forms defined in $j^s(U)$; we shall denote the forms $\phi^j, \phi^j{}_{j_1\cdots j_\nu}$, $1 \leqq \nu \leqq s-1$, and their image in $j^{s+k}(\Gamma)$ under the map $\rho_s{}^{s+k}$, by the same notation. For further reference, we give here the equations of structure satisfied by the forms $\phi^j, \phi^j{}_{j_1\cdots j_\nu}$, $1 \leqq \nu \leqq s-2$, (see [4], Lemma IV, 2):

4.2 $$\begin{aligned} d\phi^j &= \phi^j \wedge \phi^k \\ d\phi^j{}_{j_1\cdots j_\nu} &= \tfrac{1}{2} c^j{}_{j_1\cdots j_\nu i}{}^{i_1\cdots i_\sigma}{}_k{}^{k_1\cdots k_\sigma}\, \phi^i{}_{i_1\cdots i_\sigma} \wedge \phi^k{}_{k_1\cdots k_\sigma} \\ &\quad + \phi^j{}_{j_1\cdots j_\nu k} \wedge \phi^k,\ 1 \leqq \nu \leqq s-2,\ 1 \leqq j, j_1 \cdots j_\nu \leqq n, \end{aligned}$$

where $c^j{}_{j_1\cdots j_\nu i}{}^{i_1\cdots i_\sigma}{}_k{}^{k_1\cdots k_\tau}$ are constants. Let ρ_1 and ρ_2 be respectively the projections into the first and second factors of the product $j^s(U) \times j^s(U)$ and denote by Σ the closed differential system generated in $j^s(U) \times j^s(U)$ by the forms

---

[2] Observe that in [3] the map $p^s f$ is defined making $j^s f$ operate on the right; the differential forms that we use here are the image of the differential forms defined in [3] under the map $X \in j^s(U) \to X^{-1} \in j^s(V)$.

$$x^i \circ \alpha \circ \rho_1 - x^i \circ \alpha \circ \rho_2 \qquad 1 \leqq i \leqq n$$
$$\rho^*{}_1\phi^j - \rho^*{}_2\phi^j,\ \rho^*{}_1\phi^j{}_{j_1\cdots j_\nu} - \rho^*{}_2\phi^j{}_{j_1\cdots j_\nu},$$
$$1 \leqq \nu \leqq s-1,\ 1 < j, j_1 \cdots j_\nu \leqq n.$$

Put $j^s(\Gamma, U) = j^s\Gamma \cap j^s(U)$ and let $\Sigma(\Gamma, U)$ be the restriction of Σ to $j^s(\Gamma, U)$.

PROPOSITION 4.1. *A necessary and sufficient condition for a local diffeomorphism ϕ of $j^s(\Gamma, U)$ to belong to $p^s\Gamma$ is that the map $X \in U(\phi) \to (X, \phi(X))$ be a solution of $\Sigma(\Gamma, U)$.*[3]

For a proof of Proposition 4.1 see [4], Proposition IV, 6.

Take $X \in j^s(M)$, $X = (x, y, p^j{}_{j_1\cdots j_\nu})$, where $x^1, \cdots, x^n$ and $y^1, \cdots, y^n$ are coordinates around $\alpha(X)$ and $\beta(X)$ respectively. Let $\Theta = \theta^i(y) \dfrac{\partial}{\partial y^i}$ be a local vector field defined in a neighborhood of $\beta(X)$. Using Formula 1.1 and Proposition 1.1 it is easy to prove that

$$4.2 \qquad \Theta^s(X) = \sum_{i=1}^{n} \theta^i(y^1 \cdots y^n) \frac{\partial}{\partial y^i} + \sum \phi^j{}_{j_1\cdots j_\nu}\Big(p^j{}_{j_1\cdots j_k}, \frac{\partial \theta^j}{\partial y^{j_1} \cdots \partial y^{j_\nu}}\Big) \frac{\partial}{\partial p^j{}_{j_1\cdots j_\nu}}$$
$$1 \leqq k \leqq \nu,\ 1 \leqq \nu \leqq s,\ 1 \leqq j, j_1 \cdots j_\nu \leqq n.$$

Formula 4.1 shows that $j^1{}_X\Theta^s$ depends on $j^{s+1}{}_{\beta(X)}\Theta$ only. Therefore, denoting by $\tilde{T}^1{}_X$ the vector space of 1-jets of source X of vector fields defined in a neighborhood of X which are the s-th prolongation of local vector fields of M, we have a map $\eta(X) : T^{s+1}{}_{\beta(X)} \to T^1{}_X$ defined in the following way: If $\tau \in T^{s+1}{}_{\beta(X)}$ and $\tau = j^{s+1}{}_{\beta(X)}$ then $\eta(X)(\tau) = j^1{}_X\Theta^s$. By definition, $\eta(X)$ is a linear map onto $\tilde{T}^1{}_X$; since $\lambda(X)$ is injective, $\eta(X)$ is also injective.

PROPOSITION 4.2. *Let Γ be a Lie pseudo group of order s acting on M. If the dimension of $j^{s+1}{}_x\Theta(\Gamma)$ is independent of $x \in M$ then the sheaf $p^s\Theta(\Gamma)$ of germs of local vector fields of $j^s(\Gamma)$ which are the s-th prolongation of Γ vector fields is an infinitesimal Lie pseudo group of order* 1 *acting on $j^s(M)$.*

Proof. Condition 1 of Definition 3.1 clearly holds for $p^s\Theta(\Gamma)$. Since for $X \in j^s(\Gamma)$, $\eta(X)$ maps $j^{s+1}{}_{\beta(X)}\Theta(\Gamma)$ onto $j^1{}_X p^s\Theta(\Gamma)$, the dimension of $j^1{}_X p^s\Theta(\Gamma)$ is independent of $X \in j^s(\Gamma)$. To prove that condition 2 of Definition 3.1 holds for $p^s\Theta(\Gamma)$ we observe that, denoting by $\omega^j, \omega^j{}_{j_1\cdots j_\nu}$ the restrictions of the forms $\phi^j, \phi^j{}_{j_1\cdots j_\nu}$ to $j^s(\Gamma, U)$, a necessary and sufficient condition for a local diffeomorphism ϕ of $j^s(\Gamma, U)$ to belong to $p^s\Gamma$ is that ϕ leaves invariant the restrictions of the functions $x^1 \circ \alpha, \cdots, x^n \circ \alpha$ to $j^s(\Gamma, U)$ and the differential forms $\omega^j, \omega^j{}_{j_1\cdots j_\nu}$.[4] It follows that a necessary and sufficient condition for a local vector field Λ, defined in an open set V of $j^s(\Gamma, U)$, to be a section of $p^s\Theta(\Gamma)$ is that all the Lie derivatives $\mathcal{L}(\Lambda)(x^1 \circ \alpha), \cdots, \mathcal{L}(\Lambda)(x^n \circ \alpha)$ $\mathcal{L}(\Lambda)(\omega^j)$, $\mathcal{L}(\Lambda)(\omega^j{}_{j_1 \cdots j_\nu})$ vanish. But, if Λ and Θ are two vector fields, defined in the same open set V of a differentiable manifold, such that $j_x{}^1\Lambda = j_x{}^1\Theta$ for a fixed point $x \in V$, then the differential forms $\mathcal{L}(\Lambda)(\omega)$ and $\mathcal{L}(\Theta)(\omega)$ coincides at x, for any differential form ω defined in V. Hence, a necessary and sufficient condition for Λ to be a section of $p^s\Theta(\Gamma)$ is that $j^1{}_X \in j^1{}_X(p^s\Theta(\Gamma))$ for all $X \in V$.

---

[3] By definition, the map $X \in U(\phi) \to (X, \phi(X))$ is a solution of $\Sigma(\Gamma, U)$ if the image of $U(\phi)$ under this map is an integral manifold of $\Sigma(\Gamma, U)$.

[4] This is just another way of stating Proposition 4.1.

PROPOSITION 4.3. *If f and ϕ are two local diffeomorphisms of M such that for a fixed $y_0 \in M$, $j^{s+r}{}_{y_0}f = j^{s+r}{}_{y_0}\phi$ then, for any $X_0 \in j^s(M)$ with $\beta(X_0) = y_0$, $j^r{}_{X_0}(p^sf) = j^r{}_{X_0}(p^s\phi)$.*

Proof. Let (x, y, p) and (y, y', p') be two systems of coordinates of $j^s(M)$ defined respectively in the open neighborhoods U and U' of X_0 and $j^s{}_{y_0}f$. If $X = (x, y, p)$, $j^s{}_yf = (y, y'(y), p'(y))$ for $y \in U' \cap U(f)$ then, the coordinates of $(p^sf)(X)$ are

$$(x, y'(y),\ \phi^j{}_{j_1 \cdots j_\nu}(p'^j{}_{j_1 \cdots j_k}(y), p^j{}_{j_1 \cdots j_k})), \qquad 1 \leqq \nu \leqq s, 1 \leqq k \leqq \nu.$$

This shows that the jet $j^r{}_{X_0}(p^sf)$ depends only on the partial derivatives, of order at most r, of the functions $y'(y)$, $p^j{}_{j_1 \cdots j_k}(y)$.

As an easy consequence of Proposition 4.3 we have

PROPOSITION 4.4. *Let Γ be a Lie pseudo group of order s acting on M. Let f and ϕ be two elements of Γ such that, for a fixed point $y_0 \in M$, $j^{s+r}{}_{y_0}f = j_{y_0}{}^{s+r}\phi$. Denote by p^sf and $p^s\phi$ respectively the prolongation of f and ϕ to $j^s(\Gamma)$. Then, for any $X_0 \in j^s(\Gamma)$ with $\beta(X_0) = y_0$, we have $j_{X_0}{}^r(p^sf) = j_{X_0}{}^s(p^s\phi)$.*

Let Γ be a Lie pseudo group of order s and assume that Γ is also a regular and connected Lie pseudo group of order $s+1$. This means that $\Theta(\Gamma)$ is an infinitesimal Lie pseudo group of order $s+1$ and Γ is the Lie pseudo group of order $s+1$ corresponding to $\Theta(\Gamma)$. Clearly $j^s(\Gamma)$ is also connected. By Proposition 4.2 $p^s\Theta(\Gamma)$ is an infinitesimal Lie pseudo group acting on $j^s(M)$. Let Γ' be the regular and connected Lie pseudo group whose infinitesimal Lie pseudo group is $p^s\Theta(\Gamma)$ (Theorem 3.2). We shall prove that $\Gamma' = p^s\Gamma$.

PROPOSITION 4.5. *If $f \in \Gamma$, then the prolongation p^sf of f to $j^s(\Gamma)$ is in Γ'.*

Proof. Take $\phi \in \Gamma$, if ϕ belongs to a family $\phi_{1t} \circ \phi_{2t} \cdots \circ \phi_{kt}$ where $\phi_{1t}, \cdots, \phi_{kt}$ are one parameter local groups of Γ then, $p^s\phi \in \Gamma'$. Since Γ

is regular of order $s+1$ and connected, by Propositions 3.6 and 3.7 for each $x \in U(f)$ there exists $\phi \in \Gamma$ such that $j^{s+1}{}_x f = j^{s+1}{}_x \phi$ and $p^s\phi \in \Gamma'$. By Proposition 4.4 $j^1{}_X(p^s f) = j^1{}_X(p^s\phi)$ for every $X \in j^s(\Gamma)$ such that $\beta(X) = x$. Hence, $p^s f \in \Gamma'$.

PROPOSITION 4.6. $\Gamma' \subset p^s(\Gamma)$.

Proof. Let ϕ be an element of Γ' and take $X \in U(\phi)$; we may assume that $X = I^s(x)$ for some $x \in M$ for otherwise we replace ϕ by $\phi \circ p^s f$ where $f \in \Gamma$ is such that $j^s f = X$. Moreover, we may also assume that $\phi(I^s(x)) = I^s(x)$ for otherwise we replace ϕ by $(p^s f)^{-1} \circ \phi$ where $f \in \Gamma$ is such that $\phi(I^s(x)) = (p^s f)(I^s(x))$. Let U be a coordinate neighborhood of M such that $x \in U$; ϕ induces a local homeomorphism of $j^s(\Gamma, U)$ defined in an open set $V \subset U(\phi) \cap j^s(\Gamma, U)$. By Propositions 3.2 and 3.3, given $X' \in j^1(\Gamma')$ there exists $f \in \Gamma$ such that $j^1(p^s f) = X'$. Then, for each $X_0 \in V$ the jet of source X_0 of the map $X \in V \to (X, \phi(X)) \in j^s(\Gamma, U_0) \times j^s(\Gamma, U_0)$ is the jet of a solution of $\Sigma(\Gamma, U)$. Therefore, this map is itself a solution of $\Sigma(\Gamma, U)$ and by Proposition 4.1, $\phi \in p^s(\Gamma)$.

5. The equations of structure of E. Cartan.

Given a manifold M let $\rho_s^{s+k}: j^{s+k}(M) \to j^s(M)$ be the map which sends each jet $X \in j^{s+k}(M)$ into its induced jet of order s. If Γ is a Lie pseudo group acting on M we shall also denote by ρ_k^{s+k} the restriction of ρ_k^{s+k} to $j^{s+k}(\Gamma)$. Clearly ρ_k^{s+k} maps $j^{s+k}(\Gamma)$ onto $j^s(\Gamma)$.

Let M and N be two differentiable manifolds and $\rho: M \to N$ a differentiable map. We say that ρ is a fiber mapping and that (M, N, ρ) is a fibered manifold if ρ is of maximum rank at every point of M.

PROPOSITION 5.1. *Let P be a regular Lie pseudo group or order s acting on M. If Γ is also a regular Lie pseudo group of order $s+1$ then, $\rho_s^{s+1}: j^{s+1}(\Gamma) \to j^s(\Gamma)$ is a fiber mapping.*

Proof. Take $X^{s+1} \in j^{s+1}(\Gamma)$ and put $\rho_s^{s+1}(X^{s+1}) = X^s$. Let $T'_{X^{s+1}}$ be the tangent space to $A'(\alpha(X^{s+1})) = A(\alpha(X^{s+1})) \cap j^{s+1}(\Gamma)$ at the point X^{s+1}. It suffices to show that $(\rho_s^{s+1})_*$ maps $T'_{X^{s+1}}$ onto T'_{X^s}. Let $\pi_s^{s+1}: j^{s+1}\Theta(\Gamma) \to j^s\Theta(\Gamma)$ be the map which sends each element of $j^{s+1}\Theta(\Gamma)$ into the induced jet of order s. Then,

$$(\rho_s^{s+1})_* \circ \lambda^{s+1}(X^{s+1}) = \lambda^s(X^s) \circ \pi_s^{s+1}.$$

This shows that $(\rho_s^{s+1})_*$ maps $T'_{X^{s+1}}$ onto T'_{X^s}.

In the rest of this paragraph we assume that the pseudo group acting on M satisfies the following conditions:

5a. Γ is a regular Lie pseudo group of orders s and $s+1$. $j^{s+1}(\Gamma)$ is connected.

5b. For every point $x_0 \in M$ there exists a coordinate neighborhood U of x_0, with coordinates $x^1 \cdots x^n$, and an integer h satisfying the following conditions:

1) Given two points $w_1, w_2 \in U$, there exists $f \in \Gamma$ such that $f(w_1) = w_2$ if and only if $x^1(w_1) = x^1(w_2), \cdots, x^h(w_1) = x^h(w_2)$.

2) Let N be the submanifold of $U \times U$ of all couples (w_1, w_2) for which there exists $f \in \Gamma$ with $f(w_1) = w_2$. Then, $\alpha \times \beta \colon (\alpha \times \beta)^{-1}(U \times U) \subset j^s(\Gamma) \to N$ is a fiber mapping.

Remark. If Γ is transitive on M then the first part of condition 5b is always satisfied for $h = 0$. If there exists a differentiable family (f_{w_1, w_2}) of transformations of Γ defined in U and parametrized in N such that for every $(w_1, w_2) \in N$, $f_{w_1, w_2}(w_1) = w_2$, then Γ satisfies 5b.

Let $j^s(\Gamma, U, \alpha)$ be the set of all jets $X \in j^s(\Gamma, U)$ such that $x^{h+1}(\alpha(X)) = \cdots = x^n(\alpha(X)) = 0$. Clearly $j^s(\Gamma, U, \alpha)$ is a submanifold of $j^s(\Gamma, U)$ (not necessarily regularly imbedded); the maps $\beta \colon j^s(\Gamma, U, \alpha) \to U$ and $\rho_s{}^{s+1} \colon j^{s+1}(\Gamma, U, \alpha) \to j^s(\Gamma, U, \alpha)$ are fiber mappings. Denote by $(p^s\Gamma, U, \alpha)$ the restriction of $p^s\Gamma$ to $j^s(\Gamma, U, \alpha)$ and let $'\omega^j$, $'\omega^j{}_{j_1 \cdots j_\nu}$ be the restrictions of the forms ϕ^j, $\phi^j{}_{j_1 \cdots j_\nu}$ to $j^{s+1}(\Gamma, U, \alpha)$. Take a maximal set $'\omega^1 \cdots {}'\omega^{n_1}$ of linearly independent forms, at the point $I^{s+1}(x_0)$, among $'\omega^j$, $'\omega^j{}_{j_1 \cdots j_\nu}$, $1 \leqq \nu \leqq s-1$ and complete it to a maximal set $'\omega^1 \cdots {}'\omega^{n_1}$, $\varpi^1 \cdots \varpi^{n_2}$ of linearly independent forms among $'\omega^j, {}'\omega^j{}_{j_1 \cdots j_\nu}$, $1 \leqq \nu \leqq s$, at the same point. Let V^{s+1} be a neighborhood of $I^{s+1}(x)$ in the manifold $j^{s+1}(\Gamma, U, \alpha)$ where these forms remain linearly independent. By formulas 4.2, we have in V^{s+1} the following equations

$$d'\omega^i = \tfrac{1}{2} c^i{}_{jk}{}'\omega^j \wedge {}'\omega^k + a^i{}_{j\lambda}{}'\omega^j \wedge {}'\varpi^\lambda$$

where $c^i{}_{jk}$ and $a^i{}_{j\lambda}$ are functions defined in V^{s+1} invariant under the restriction $(p^{s+1}\Gamma, U, \alpha) \mid V^{s+1}\Gamma, U, \alpha)$ to V^{s+1}. Let V^s be a sufficiently small neighborhood of $I^0(x)$ in $j^s(\Gamma, U, \alpha)$ and $\tau \colon V^s \to V^{s+1}$ a cross section with respect to $\rho_s{}^{s+1}$. Put $\omega^i = (\rho_s{}^{s+1})^* {}'\omega^i$, $\varpi^\lambda = (\rho_s{}^{s+1})^* {}'\varpi^\lambda$. By formulas 4.1, at each point of V^{s+1}, the forms $'\omega^i$, $'\omega^\lambda$ generate the space of linear forms which vanish on the vectors tangent to the fibre of $\rho_s{}^{s+1} \colon j^{s+1}(\Gamma, U, \alpha) \to j^s(\Gamma, U, \alpha)$. Hence the forms ω^i, ϖ^λ form a basis of the space of linear differential forms on V^s. Moreover,

$$d\omega^i = \tfrac{1}{2} c^i{}_{jk} \omega^j \wedge \omega^k + a^i{}_{j\lambda} \omega^j \wedge \varpi^\lambda,$$

where c^i_{jk}, $a^i_{j\lambda}$ are functions defined in V^s invariant under $(p^s\Gamma, U, \alpha)V^s$. At each point of V^s the forms $\omega^1 \cdots \omega^{n_1}$ generate the same vector space as the restrictions of the forms $\phi^j, \phi^j_{j_1 \cdots j_\nu}$, $1 \leqq \nu \leqq s-1$, to V^s. By Proposition IV, 8 of [3], a necessary and sufficient condition for a local diffeomorphism f of V^s to belong to $(p^s\Gamma, U, \alpha) \mid V^s$ is that f leaves invariant the restrictions of the functions $x^1 \circ \alpha \cdots x^h \circ \alpha$ to V^s and the forms $\omega^1 \cdots \omega^{n_1}$. We have proved the following theorem:

THEOREM 5.1. *There exists a neighborhood V^s of $I^s(x_0)$ in $j^s(\Gamma, U, \alpha)$, functions $g_1 \cdots g_h$ and differential forms $\omega^1 \cdots \omega^{n_1}, \varpi^1 \cdots \varpi^{n_2}$* defined in V^s such that:

1) A local diffeomorphism f of V^s belongs to $(p^s\Gamma, U, \alpha) \mid V^s$ if and only if f leaves invariant the functions $g^1 \cdots g^h$ and the forms $\omega^1 \cdots \omega^{n_1}$.

2) The differentials $dg^1 \cdots dg^h$ are linearly independent in V^s. The differentials $\omega^1 \cdots \omega^{n_1}$, $\varpi^1 \cdots \varpi^{n_2}$ are linearly independent in V^s.

3) $d\omega^i = \frac{1}{2} c^i_{jk} \omega^j \wedge \omega^k + a^i_{j\lambda} \omega^j \wedge \varpi^\lambda$
where $c^i_{j\lambda}$ and a^i_{jk} are functions defined in V^s invariant under $(p^s\Gamma, U, \alpha) \mid V^s$.

Definition 5.1. A regular Lie pseudo group Γ acting on a manifold M is a Cartan pseudo group if:

1) There exist functions $g_1 \cdots g_h$ and differential forms $\omega^1 \cdots \omega^{n_1}$, $\varpi^1 \cdots \varpi^{n_2}$ definde in M such that a local differomorphism f of M is in Γ if and only if f leaves invariant the functions $g_1 \cdots g_h$ and the forms $\omega^1 \cdots \omega^{n_1}$.

2) The differentials $dg_1 \cdots dg_h$ are linearly independent and the forms $\omega^1 \cdots \omega^{n_1}, \varpi^1 \cdots \varpi^{n_2}$ form a basis of the space of linear forms at every point of M.

3) $d\omega^i = c^i_{jk} \omega^j \wedge \omega^k + a^i_{j\lambda} \omega^j \wedge \varpi^\lambda$ where the c^i_{jk}, $a^i_{j\lambda}$ are functions defined in M invariant under the transformations of Γ.

Definition 5.2. Let (M, N, ρ) be a fibered manifold and let $\bar{f}$, f be two local diffeomorphisms of M and N respectively. We say that $\bar{f}$ is a prolongation of f if $\rho(U(\bar{f})) = U(f)$ and $f \circ \rho = \rho \circ \bar{f}$.

Summing up the results of this paragraph we have the following theorem:

THEOREM 5.2. *Let Γ be a Lie pseudo group acting on M and satisfying the conditions* 5a *and* 5b. *Then, for every point $x_0 \in M$ there exists a neighborhood U of x_0, a fibered manifold $(\tilde{U}, U, \rho)$ and a Cartan pseudo group $\tilde{\Gamma}$ acting in $\tilde{U}$ with the following properties.*

1) *For each* $x \in U$ *there exists a neighborhood* V *of* $I^0(x) \subset j^s(\Gamma)$ *such that, for every* $f \in \Gamma$ *with* $j_y^s f \in V$ *for all* $y \in U(f)$, *there exists* $f \in \tilde{\Gamma}$ *which is a prolongation of* f.

2) *Given* $\tilde{f} \in \tilde{\Gamma}$, *for each* $\tilde{x} \in U(\tilde{f})$ *there exists a neighborhood* $\tilde{V}$ *of* $\tilde{x}$ *such that* $\tilde{f} \mid \tilde{V}$ *is a prolongation of an element* $f \in \Gamma$.

3) *If* $\tilde{f}_1 \tilde{f}_2 \in \tilde{\Gamma}$, $U(\tilde{f}_1) = U(\tilde{f}_2)$ and $\tilde{f}_1$ and $\tilde{f}_2$ *are the prolongation of the same element of* Γ *then* $\tilde{f}_1 = \tilde{f}_2$.

INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] E. Cartan, "Sur la structure des groupes infinis de transformations," *Ann. Ec. Norm.*, tome 21, 1904.

[2] C. Ehresmann, "Introduction a la theorie des structures infinitesimales et des pseudo-groupes de Lie," *Colloque de Geometrie Differentielle*, Strasbourg, 1953, 97-117.

[3] ———, *Categories topologiques et categories differentiables, Colloque de Geometrie Differentielle Globale, Brussels*, 1958.

[4] M. Kuranishi, "On the local theory of continuous infinite pseudo groups II," *Nagoya Mathematical Journal*, vol. 19, 1961.

[5] P. Libermann, "Pseudogroupes infinitesimaux attaches aux pseudogroupes de Lie," *Bulletin de la Société Mathematique de France*, tome 87 (1959), pp. 409-425.

[6] S. Lie, *Die Grundlagen für die Theorie der unendlichen kontinuierlichen Transformationsgruppen I, II, Gesammelte Abhandlungen*, Vol. 6, 300-364.

[7] K. Nomizu, *Lie groups and differential geometry*, Publications of the Mathematical Society of Japan, 1956.

[8] R. Palais, "A global formulation of the Lie theory of transformations groups," *Memoirs of the American Mathematical Society*, number 22.

CUSP FORMS ARISING FROM HILBERT'S MODULAR FUNCTIONS FOR THE FIELD OF $3^{\frac{1}{2}}$.*[1]

By HARVEY COHN.[2]

---

27. Introduction. The first two parts ([16] and [25]) of this series show how the decomposition into four and three integral squares is as easy a problem in the field of $2^{\frac{1}{2}}$ as it was in the field of $5^{\frac{1}{2}}$, (see [5] and [12]). We also saw that for the field of $3^{\frac{1}{2}}$, the results are "almost" as simple, but are obscured by the presence of unknown cusp forms. In this paper, we shall see that the cusp forms thus arising (for four squares) for the field of $3^{\frac{1}{2}}$ are really a blessing, in that they make for a particularly interesting modular transformation theory. This is perhaps consistent with the apparent paradox that Hecke's pioneering work on class number [17] reveals *formal simplifications* in $R(3^{\frac{1}{2}})$ not present in the other fields.

We begin by summarizing the decomposition formulae for an alegbraic integer ν as the sum of four squares in $R(D^{\frac{1}{2}})$ for $D = 1, 5, 8$, and 12. We count the number of ordered triples $(\xi_1, \xi_2, \xi_3, \xi_4)$ whose elements are integers in $R(D^{\frac{1}{2}})$ for which

$$\nu = \xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2. \tag{27.1}$$

It is necessary that ν be totally positive, and when $D = 8$ or 12, it is necessary that ν also be of the form $a + bD^{\frac{1}{2}}$ (a and b rational integers). Now we let $A_4(\nu)$ denote the number of such decompositions, (with D understood), and we let

$$\begin{cases} D(\mu) = \sum |N(\nu)|, & \mu/\nu = \text{integer} \\ D^*(\mu) = \sum |N(\nu)|, & \mu/\nu = \text{odd integer}, \end{cases} \tag{27.2}$$

where the summations are over the (ideal) divisors of μ as indicated and an odd integer is one of odd norm; e.g., here $N(\mu/\nu)$ odd. (There are obvious

---

* Received August 8, 1961.

[1] Work supported by Research Grant G-7412 of the National Science Foundation. Abstract appeared in *Notices of the American Mathematical Society*, vol. 8 (1961), p. 55.

[2] The present work is a continuation of items [16] and [25] cited among the references below. Sections and bibliographical items are numbered consecutively, §§ 1-11 and references [1]-[15] appear in item [16] while §§ 12-26 and references [16]-[23] appear in item [25]. Notations are generally preserved with the main exception that we use $\Theta(d, c; \tau)$ of [25, § 14] in preference to $\Theta_{d,c}(\tau)$ of [16, § 3], (with the roles of c and d interchanged).

specializations of course when $D = 1$, and $\nu = N(\nu)$ is a positive rational integer while the μ are just positive divisors, etc.) The previously known results ([4], [5], [16]) are given in the following table:

FORMULAS FOR $A_4(\mu)$

D	2^t	$A_4(\mu)$ if $2^t \nmid N(\mu)$	$A_4(\mu)$ if $2^t \mid N(\mu)$
1 (Jacobi)	2	$8D(\mu)$	$48D^*(\mu) - 24D(\mu)$
5 (Götzky)	4	$8D(\mu)$	$120D(\mu) - 144D^*(\mu)$
8	8	$8D^*(\mu)$	$48D(\mu) - 84D^*(\mu)$
12	8	$4D^*(\mu) + 4L(\mu)$	$24D(\mu) - 42D^*(\mu)$

Here $4L(\mu)$ is the error function (called $E(\mu)$ in [16]), which will be shown in §§ 36-38 (below) to vanish if $8 \mid N(\mu)$. The formulas above can be put in many different forms but the ones above are believed to be the most convenient for all four values of D simultaneously.

The error function $4L(\mu)$ is the most immediate object of our investigation. We shall show that $L(\mu)$ is very much like the Ramanujan function $\tau(n)$ with corresponding multiplicative properties, proved in a manner reminiscent of Mordell's method [31], even complete with a corresponding Hecke-transformation theory [29], (see §§ 39-40 below). More significant, however, is the fact that $L(\mu) = 0$ whenever certain types of prime power factors are present in μ. These results are contained in Theorem 34. 1 (below). The corresponding modular-functional identities, which arise in the course of these theorems, are also of value in that they give rise to a set of modular forms of dimension -2 in two complex variables, (denoted by $M(\tau, \tau')$, $P(\tau, \tau')$ and $K(\tau, \tau')$ in § 37 below), whose ratios display properties like the ordinary modular invariants and which are probably the simplest case of Hecke's bi-quadratic complex multiplication theory [17].

A fruitful development has been joint work suggested by G. Pall [26] using the reduction of the representation problem (27.1) to a completely rational problem in quaternary forms (with a side condition). There it is shown directly that in $R(12^{\frac{1}{2}})$, $\mu = a + b \cdot 12^{\frac{1}{2}} \gg 0$ is always representable as the sum of four squares, (which has not yet been completely proved from modular functions theory alone, particularly for μ odd). Here the remark should be made that the results of Götzky and Maass for $R(5^{\frac{1}{2}})$ had not been extended earlier to other fields because when $D \neq 5$ one does not know all the classes in the genus of $\xi_1^2 + \xi_2^2 + \xi_3^2 + \xi_4^2$ or $\xi_1^2 + \xi_2^2 + \xi_3^2$ for $R(D^{\frac{1}{2}})$, as required by Siegel's method [15]. Although this difficulty was circumvented by constructing modular invariants in [16] and [25], nevertheless when a reduction to *rational* quarternary forms is made, a well developed genus theory is possible, which even promises to account for the interesting behavior of $L(\mu)$, (of § 41 below). The present paper continues to develop modular functions of two complex variables as the principal tool.

28. Notation. Preserving the earlier notation we let τ and τ' be two independent complex variables with $\operatorname{Im}\tau > 0$, $\operatorname{Im}\tau' < 0$. We let lower case Greek letters generally denote integers in the field $R(D^{\frac{1}{2}})$ and we let roman letters denote rational integers. In most of the paper $D = 12$, but we shall occasionally state results for $D = 8$. In either case we let ϵ denote the fundamental unit and η denote a prime divisor of 2 in the following fashion:

$$(28.1) \qquad \left\{ \begin{array}{rlrl} \epsilon &= 2 + 3^{\frac{1}{2}}, & \epsilon &= 1 + 2^{\frac{1}{2}} \\ \eta &= 1 + 3^{\frac{1}{2}}, & \eta &= 2^{\frac{1}{2}} \\ N(\epsilon) &= +1, & N(\epsilon) &= -1 \\ \eta^2 &= 2\epsilon, & \eta^2 &= 2 \\ \epsilon &= 1 + \eta, & \epsilon &= 1 + \eta \end{array} \right.$$

As before, we define norm N and trace S to act on conjugates in $R(D^{\frac{1}{2}})$ (denoted by primes) and on τ and τ' in like fashion. We do not always show both variables if like operations are being performed on each. We use the rings

$$(28.2) \qquad \mathfrak{O} = [1, \epsilon] = [1, \eta]$$

and

$$(28.3) \qquad \mathfrak{O}_2 = [1, 2\eta] = [1, D^{\frac{1}{2}}].$$

We also let $\mathfrak{O}^+$ and $\mathfrak{O}_2^+$ denote the totally positive integers in each. The functions are written as Fourier Series in terms of

$$(28.4) \qquad \begin{aligned} e(\zeta\tau) &= \exp \pi i S(\zeta\tau/D^{\frac{1}{2}}) \\ &= \exp \pi i(\zeta\tau - \zeta'\tau')/D^{\frac{1}{2}}. \end{aligned}$$

Thus our basic interest is in

$$(28.5) \qquad \Theta(d, c; \tau) = \sum_{\nu \in \mathfrak{O}} e(\nu d\eta + [\nu + c\eta/2]^2 \tau)$$

and the earlier papers ([16, § 6] and [25, § 21]) considered the "singular" series (for $D = 12$)

$$\Omega_{00}(\tau) = \sum_{\mu \in \mathfrak{O}_2^+} B^*(\mu) e(\mu\tau) \tag{28.6}$$

where

$$B^*(\mu) = \begin{cases} 4D^*(\mu), & 8 \nmid N(\mu) \\ 24D(\mu) - 42D^*(\mu), & 8 \mid N(\mu) \end{cases} \tag{28.7}$$

Then $\Theta(0,0;\tau)^4 - \Omega_{00}(\tau)$ equals

$$4\Lambda(\tau) = 4 \sum_{\mu \in \mathfrak{O}_2^+} L(\mu) e(\mu\tau) \tag{28.8}$$

where

$$L(\mu) = A_4(\mu) - B^*(\mu). \tag{28.9}$$

For $D = 8$, more simply, Θ^4 equals the corresponding singular series $\Omega_{00}(\tau)$.

29. Definition of some groups. We are interested in transformations from (τ, τ') to (τ_1, τ_1')

$$\tau_1 = \frac{\alpha\tau + \beta}{\gamma\tau + \delta}, \qquad \tau_1' = \frac{\alpha'\tau' + \beta'}{\gamma'\tau' + \delta'} \tag{29.1}$$

subject to the condition that $\alpha\delta - \beta\gamma$ is a totally positive unit. In using matrices we make the convention that the matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ will denote the aggregate $\omega \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ where ω is a unit. Capital German letters will be used for groups and small German letters for elements. The identity will be denoted by $\mathfrak{e}$.

We consider the following groups:

$\mathfrak{O}$ the complete Hilbert group for which $\alpha\delta - \beta\gamma$ is a totally positive unit,

$\mathfrak{G}$ the sub-group of $\mathfrak{O}$ for which $\alpha\delta - \beta\gamma$ is the square of a unit,

$\mathfrak{H}$ the sub-group of $\mathfrak{G}$ for which the elements α, β, γ, δ lie in $\mathfrak{O}_2$ (using an appropriate unit factor ω),

$\mathfrak{K}$ the sub-group of $\mathfrak{H}$ for which

$$\omega \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \pmod{2},$$

$\mathfrak{J}$ the sub-group of $\mathfrak{K}$ for which

$$\omega \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \equiv \begin{pmatrix} \pm\epsilon^{2t} & 2\rho \\ 2\sigma & 1 \end{pmatrix} \quad \text{or} \begin{pmatrix} 2\rho & \pm\epsilon^{2t} \\ 1 & 2\sigma \end{pmatrix} \pmod{4}$$

for some choice of sign, $t = 0$ or 1, and ρ, σ in $\mathfrak{O}$.

THEOREM 29.1. *The following coset properties are valid:*

		$D=8$	$D=12$
(a)	$\mathfrak{O}=$	$\mathfrak{G}$	$\mathfrak{G}\left\{e+\begin{pmatrix}\epsilon & 0\\ 0 & 1\end{pmatrix}\right\}$
(b)	$\mathfrak{G}=$	$\mathfrak{H}\left\{e+\begin{pmatrix}1 & \eta\\ 0 & 1\end{pmatrix}+\begin{pmatrix}0 & -1\\ 1 & \eta\end{pmatrix}+\begin{pmatrix}1 & \eta\\ 1 & \epsilon\end{pmatrix}\right\}$	
(c)	$\mathfrak{H}=$	$\mathfrak{K}\left\{e+\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix}+\begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}\right\}$	
(d)	$\mathfrak{K}=$	$\mathfrak{I}$	$\mathfrak{I}\left\{e+\begin{pmatrix}1+2\eta & 2\eta\\ -2\eta & 1-2\eta\end{pmatrix}\right\}$

Proof. The result for (a) is obvious. The result for (b) was proved in [16, § 3] in another form, but we might first note that every group element of $\mathfrak{O}$ is congruent $(\operatorname{mod}\eta)$ to one of the elements

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix},\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix},\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}. \tag{29.2}$$

An examination of the individual cases modulo 2 can then give the result directly, (since the factor $\omega=\epsilon\equiv 1+\eta \bmod 2$). The result for (c) is evident from the matrices (29.2), taken modulo 2 this time, since every integer of $\mathfrak{O}_2$ is congruent to a rational integer.

The last result, for (d), is shown by noticing that $\epsilon^2\equiv -1+2\eta\ (\operatorname{mod} 4)$ when $D=8$ while $\epsilon^2\equiv -1\ (\operatorname{mod} 4)$ when $D=12$. Concentrating on the latter case, we see if $\alpha\equiv 1\ (\operatorname{mod} 2)$ but $\alpha\not\equiv 1\ (\operatorname{mod} 2\eta)$ then $-\alpha\equiv\epsilon^2\alpha\equiv 1\ (\operatorname{mod} 2\eta)$. Thus any group element of $\mathfrak{K}$, by proper choice of ω, corresponds to the matrices

$$\begin{pmatrix}1+2\eta\alpha_1 & 2\beta_1\\ 2\gamma_1 & 1+2\eta\delta_1\end{pmatrix} \text{ or } \begin{pmatrix}2\beta_1 & 1+2\eta\alpha_1\\ 1+2\eta\delta_1 & 2\gamma_1\end{pmatrix}$$

for properly chosen α_1, β_1, γ_1, δ_1 in $\mathfrak{O}$. By the determinant condition, however $\alpha_1\equiv\delta_1\ (\operatorname{mod}\eta)$, hence the two cosets occur depending on whether $\alpha_1\equiv\delta_1\equiv 0$ or 1.

Q.E.D.

The groups shown indicate that "$D=12$" is richer in subgroups than "$D=8$," (actually to our advantage). The invariance properties leading to these sub-groups can be derived independently of modular functions by recalling that every fraction α/β in $R(D^{\frac{1}{2}})$ is congruent modulo 2 to one of the six quantities

$$0, 1, \infty, \epsilon, \eta, 1/\eta. \tag{29.3}$$

Here $\infty = 1/0$ and $1/\eta$ are both "infinite" elements and the rational operations can be performed as usual [16, § 6] (barring indeterminacies). We note the (unordered) triples:

$$[00] = \{0, 1, \infty\}, \qquad [01] = \{0, \epsilon, 1/\eta\}$$
$$[10] = \{\epsilon, \eta, \infty\}, \qquad [11] = \{1, \eta, 1/\eta\}. \tag{29.4}$$

THEOREM 29.2. *The group* $\mathfrak{H}$ *can be defined as the sub-group of* $\mathfrak{G}$ *which preserves the triple* [00]. *The cosets of* $\mathfrak{G}/\mathfrak{H}$ *transform* [00] *into* [00], [10], [01], *or* [11] *in the order shown in Theorem* 29.1(b) *and the group* $\mathfrak{K}$ *can be defined as the sub-group of* $\mathfrak{H}$ *which preserves the triple* [11].

These properties of $\mathfrak{G}$, $\mathfrak{H}$, and $\mathfrak{K}$ can be proved by inspection. We make no effort, however, to be equally elegant with $\mathfrak{J}$.

THEOREM 29.3. *If we define* $\mathfrak{E}_4$ *as the sub-group of* $\mathfrak{G}$ *whose elements are* $\equiv e \pmod 4$ *then* $\mathfrak{J}$ *is generated by*

$$\mathfrak{E}_4, \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2\eta \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} \delta^2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{29.5}$$

The proof, again, is detailed but straight-forward. For example, working with the generators listed above, we can construct

$$\begin{pmatrix} 1 & 2\alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2\beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -2\alpha & 1-4\alpha\beta \\ -1 & -2\beta \end{pmatrix}, \text{ etc.}$$

We do not make any effort to explore other quotients (such as $\mathfrak{J}/\mathfrak{E}_4$) since they are not directly required for the ensuing analysis.

30. Generators. It was noted in [16, § 2], by virtue of the Euclidean algorithm, that sufficient generators of $\mathfrak{G}$ are

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} \epsilon^3 & 0 \\ 0 & 1 \end{pmatrix}. \tag{30.1}$$

In general, the problem of verifying a set of generators for an arbitrary sub-group of $\mathfrak{G}$ is quite difficult. We might easily conjecture that a sufficient set of generators of $\mathfrak{H}$ is the following:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2\eta \\ 0 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} \epsilon^2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{30.2}$$

but it is not clear that this result is true. Let us call $\mathfrak{H}^*$ the group generated by the elements (30.2). We can prove the following result which is almost as useful as knowing if $\mathfrak{H}^*$ equals $\mathfrak{H}$:

LEMMA 30.1. *If*

$$(30.3) \qquad \mathfrak{s}_1 = \begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix}, \quad \mathfrak{s}_2 = \begin{pmatrix} 0 & -1 \\ 1 & \eta \end{pmatrix}, \quad \mathfrak{s}_3 = \begin{pmatrix} 1 & \eta \\ 1 & \epsilon \end{pmatrix}$$

are the coset representatives in Theorem 29.1(b) and if $\mathfrak{g}_j$ is any one of the four elements in (30.1) then the 12 elements $\mathfrak{s}_i\mathfrak{g}_j$ each determine an $\mathfrak{s}_k$ such that $\mathfrak{s}_i\mathfrak{g}_j\mathfrak{s}_k^{-1} = \mathfrak{h}^*$ lies in $\mathfrak{H}^*$.

This result is proved by actual trial and error, and it is needed later on.

It would also seem plausible that $\mathfrak{J}$ is generated by the last four matrices in (29.5), (which would generate $\mathfrak{E}_4$), but, again, a general device for verifying generators is beyond the scope of this work.

31. Functions defined on a group. We shall be dealing with functions (or forms) defined on a sub-group $\mathfrak{M}$ of $\mathfrak{Q}$ and of negative even dimension $-2w$. These are functions of τ and τ' (but abbreviated as functions of τ) for which

$$(31.1) \qquad F(\mathfrak{h}\tau) = F(\tau)/N(\text{den}\,\mathfrak{h}\tau)^{2w}$$

where $\mathfrak{h}$ is a transformation in $\mathfrak{M}$. Here

$$(31.2) \qquad \mathfrak{h}\tau = (\alpha\tau + \beta)/(\gamma\tau + \delta),$$

and the "norm of denominator" means

$$(31.3) \qquad N(\text{den}\,\mathfrak{h}\tau) = N(\gamma\tau + \delta) = (\gamma\tau + \delta)(\gamma'\tau' + \delta').$$

We make the usual restriction that $F(\tau, \tau')$ have no finite singularities and that $F(\tau)/N(\gamma\tau + \delta)^{2w}$ approach a limit as $\gamma\tau + \delta$ and $\gamma'\tau' + \delta'$ approach 0 vertically for $\gamma \neq 0$, and, (when $\gamma = 0$), $F(\tau)$ approach a limit as τ and τ' approach ∞ vertically. If these limits are zero we call $F(\tau)$ a cusp function (or cusp form).

THEOREM 31.1. *If $F(\tau)$ is defined on $\mathfrak{H}^*$ then we can define the following four functions:*

$$(31.4) \qquad \begin{aligned} F(0,0;\tau) &= F(\tau) \\ F(1,0;\tau) &= F(\tau + \eta) \\ F(0,1;\tau) &= F(-1/[\tau + \eta])/N(\tau + \eta)^{2w} \\ F(1,1;\tau) &= F([\tau + \eta]/[\tau + \epsilon])/N(\tau + \epsilon)^{2w}. \end{aligned}$$

These functions transform under $\mathfrak{G}$ *according as the triples* (29.4), *and* $F(\tau)$ *is defined on* $\mathfrak{H}$.

Proof. For convenience, let us omit mention of the norms of denominators, which trivially take care of themselves. The functions in (31.4) are of the form $F(\mathfrak{s}_i\tau)$; we apply the substitution of $\mathfrak{g}\tau$ for τ to see how $F(\mathfrak{s}_i\tau)$ transforms, (where $\mathfrak{s}_i$ refers to (30.3) and $\mathfrak{g}$ refers to (30.1)). Now, by Lemma 30.1, $\mathfrak{s}_i\mathfrak{g} = \mathfrak{h}^*\mathfrak{s}_k$, thus

$$F(\mathfrak{s}_i\mathfrak{g}\tau) = F(\mathfrak{h}^*\mathfrak{s}_k\tau) = F(\mathfrak{s}_k\tau) \tag{31.5}$$

by the invariance of F on $\mathfrak{H}^*$. Since $\mathfrak{H}^* \subset \mathfrak{H}$ then $\mathfrak{s}_i\mathfrak{g}\mathfrak{s}_k^{-1} = \mathfrak{h}^* \in \mathfrak{H}$ and F transforms as though it belonged to $\mathfrak{H}$. Thus Theorem 29.2 implies (31.4).

Q. E. D.

We now recall that $\Theta^4(\tau) = \Theta(0,0,\tau)^4$ is defined on $\mathfrak{H}$ with $w = 1$ and leads to the set (28.5).

We recall, from [16, § 6], that a function $\Psi_0(2\tau) = \Omega_8(\tau)$ exists which is defined not only on $\mathfrak{H}$ but on $\mathfrak{G}$ (in fact on $\mathfrak{Q}$); while the singular series $\Omega_{0,0}(\tau)$ of [16, § 6] was likewise defined on $\mathfrak{H}$. (These functions need not be reproduced here.) We shall need the transformation equations of $F(\tau) = F(0,0;\tau)$ a function belonging to $\mathfrak{H}$:

$$\left\{\begin{aligned} F(d,c;-1/\tau) &= F(c,d;\tau)N(\tau)^2 \\ F(d,c;\epsilon^2\tau) &= F(d,c;\tau) \\ F(d,c;\tau+1) &= F(d+c,c;\tau) \\ F(d,c;\tau+\eta) &= F(d+1,c;\tau) \end{aligned}\right. \tag{31.6}$$

where d and c are determined modulo 2.

In concluding this section, let us note that the behavior of the denominators ignored in the proof of Theorem 31.1 can be explained by noting the invariance of $F(\tau)/\Psi_0(2\tau)^w$ or, even more simply, the invariance of $F(\tau,\tau')\,(d\tau d\tau')^w$.

32. Conditions for determining functions on $\mathfrak{H}$. In [16, § 9] we saw that the only functions of dimension -2 defined on $\mathfrak{H}$ for $D = 8$ are those mentioned above, namely $\Theta(0,0;\tau)^4 = \Omega_{00}(\tau)$ and $\Psi_0(2\tau)$.

For the remainder of this paper we consider only $D = 12$.

To accomodate the case of $F(\tau)$ defined on $\mathfrak{H}$ and with dimension $-2w \neq -2$, we must repeat part of the procedure of [16, § 10]. Make the transformation

(32.1) $$\tau = u + v, \quad \tau' = -u + v.$$

Then define

(32.2) $$F^{(t)}(u) = (\partial/\partial v)^t F(1,1;\tau)\,|_{v=0}.$$

It was seen that if we can show

(32.3) $$F^{(0)}(u) \equiv F^{(1)}(u) \equiv \cdots \equiv F^{(t-1)}(u) \equiv 0,$$

then for $\operatorname{Re} u > 0$

(32.4) $$\begin{cases} F^{(t)}(u + 3^{\frac{1}{2}}) = F^{(t)}(u) \\ F^{(t)}(-1/u) = F^{(t)}(u)\,u^{4w+2t}. \end{cases}$$

For a cusp form, of course, $F^{(t)}(u) \to 0$ as $\operatorname{Im} u \to \infty$.

Actually $F(1,1;\tau)$ has a series expansion

(32.5) $$F(1,1;\tau) = \sum_{\mu \in \mathfrak{O}^+} A_\mu e(\mu\tau),$$

since

$$F(1,1;\tau + 2) = F(1,1;\tau + 2\eta) = F(1,1;\tau).$$

If we substitute (32.1) together with

(32.6) $$\mu = a + b3^{\frac{1}{2}},$$

then we obtain, by a brief calculation,

(32.7) $$F^{(t)}(u) = \sum_{a=1}^{\infty} \exp \pi i a u/3^{\frac{1}{2}} \sum_{\mu \in \mathfrak{O}^+} A_\mu (\pi i b)^t,$$

(under the convention $0^0 = 1$, for $t = 0$).

We recall an earlier definition of a *symmetric* function as one for which $F(\tau)$ is unchanged under the change of sign of the radical $3^{\frac{1}{2}}$; an equivalent and perhaps more useful form is $F(\tau, \tau') = F(-\tau', -\tau)$. It is easily seen that the symmetry carries over to all $F(d, c; \tau)$ since the conjugates of $\mathfrak{S}_t$ in (30.3) determine the same cosets. But then $A_\mu = A_{\mu'}$ and $+b$ is balanced by $-b$ in (32.7). *Hence for symmetric* $F(\tau), F^{(t)}(u) \equiv 0$ *for odd* t.

Now by the method of [16, § 10] it suffices to know that

(32.8) $$F^{(0)}(u) \equiv F^{(1)}(u) \equiv \cdots \equiv F^{(4w-1)}(u) \equiv 0$$

in order to assert that $F(\tau) \equiv 0$. We recall that the proof consists of knowing $\Theta(1,1;\tau)$ vanishes only on the manifold $v = 0$ (called the "zero manifold") and on images of this manifold under $\mathfrak{G}$ (with simple zeros in v).

When $w = 1$, we saw for *symmetric* F the only solution to systems (32.4) for $0 \leqq t \leqq 3$ occurs when $t = 2$ (see [16, § 10]), and it accounts for the cusp form $\Lambda(\tau)$ as the only other function defined on $\mathfrak{H}$ to within a constant factor, in addition to those mentioned in § 31 (above).

We note more generally that a cusp-form solution to (32.4) is of the type

$$F^{(t)}(u) = [I'(u)]^{t+2w} f_s(I(u)) I(u)^{-s_1}[I(u)-1]^{-s_2} \tag{32.8}$$

where $I(u)$ is Hecke's modular invariant [16, § 10] and $s_1 = w + [t/2]$, $s_2 = [(5t+10w)/6]$. If the degree of the polynomial f_s is s, then, since $I(u)$ behaves like $\exp -2\pi iu/3^{\frac{1}{2}}$ as $\operatorname{Im} u \to \infty$, it follows that

$$F^{(t)}(u) = \sum_{n=s_0}^{\infty} c_n \exp 2\pi inu/3^{\frac{1}{2}}, \qquad c_{s_0} \neq 0, \tag{32.9}$$

where

$$s_0 = [t/2] + [(5t+10w)/6] - w - t - s \leqq t/3 + 2w/3 - s. \tag{32.10}$$

Since $t \leqq 4w - 1$, if $s_0 \geqq 2w$ then $s < 0$ and the polynomial f_s must vanish. We summarize our results as follows:

THEOREM 32.1. *The only symmetric cusp forms belonging to* $\mathfrak{H}$ *and of dimension* -2 *are constant multiples of* $\Lambda(\tau)$.

THEOREM 32.2. *A cusp form belonging to* $\mathfrak{H}$ *and of dimension* $-2w$ *vanishes identically if the expansion* (32.7) *begins with a term of order* $a \geqq 4w$.

33. The types of primes. To understand the coefficient function $L(\mu)$ we must first consider the factorization of an arbitrary integer μ of $\mathfrak{O}$, (always for $R(3^{\frac{1}{2}})$). The rational primes 2, 3, q, p, r lead to primes in $\mathfrak{O}$ in the following manner:

$$2 = (3^{\frac{1}{2}}+1)^2/(2+3^{\frac{1}{2}}) = \eta^2/\epsilon$$
$$3 = (3^{\frac{1}{2}})^2, \ (3^{\frac{1}{2}}\epsilon = 3 + 2\cdot 3^{\frac{1}{2}} \in \mathfrak{O}_2),$$

If $q \equiv \pm 5 \pmod{12}$, q is unfactorable.

If $p \equiv 1 \pmod{12}$, we can uniquely represent $p = a^2 - 12b^2$ or $p = \pi\pi'$ where, (choosing $a > 0, b > 0$),

$$\pi = a + 2\cdot 3^{\frac{1}{2}}b \in \mathfrak{O}_2^{+} \tag{33.1}$$

If $r \equiv -1 \pmod{12}$, we can uniquely represent $-r = a^2 - 12b^2$, (but can not so represent r). Again, (choosing $a > 0, b > 0$), $r = -\rho\rho'$, where

$$\rho = a + 2b\cdot 3^{\frac{1}{2}} \in \mathfrak{O}_2 \text{ but } \notin \mathfrak{O}_2^{+}. \tag{33.2}$$

Thus we see the prime divisors of 2, 3, r (as well as their associates) can not belong to $\mathfrak{O}_2^{+}$ while prime divisors of p and q can be chosen to belong to $\mathfrak{O}_2^+$. We shall continue to use the symbols p, q, and r in the current meaning.

It is especially significant that we can have $\mu = \mu_1\mu_2$ where $(\mu_1, \mu_2) = 1$ and $\mu \in \mathfrak{O}_2^+$ while μ_1 and μ_2 can belong just to $\mathfrak{O}^+$ (and not $\mathfrak{O}_2^+$).

34. Statement of multiplication theorem. The theorem for expressing $L(\mu_1\mu_2)$ in terms of $L(\mu_1)$ and $L(\mu_2)$ is quite elaborate. In preparation, recall from § 28 above that for $\mu \in \mathfrak{O}_2^+$

$$4L(\mu) = A_4(\mu) - B^*(\mu). \tag{34.1}$$

We verify, to start, that $L(\mu\epsilon^2) = L(\mu)$, and,

$$L(1) = 1, \qquad L(2) = 2, \qquad L(2\epsilon) = -2. \tag{34.2}$$

THEOREM 34. 1. *Let* $\mu \in \mathfrak{O}_2^+$ *and let*

$$\mu = \eta^s \rho \qquad \rho \textit{ is odd.} \tag{34.3}$$

Then $s > 1$ *necessarily and, either* ρ *or* $\rho\epsilon$ *or both* $\in \mathfrak{O}_2^+$. *Thus*

$$L(\mu) = \begin{cases} 2L(\rho) \textit{ if } \rho \in \mathfrak{O}_2^+,\ s = 2, \\ -2L(\rho\epsilon) \textit{ if } \rho\epsilon \in \mathfrak{O}_2^+,\ s = 2, \\ 0 \textit{ if } s \geqq 3. \end{cases} \tag{34.4}$$

Next factor an odd $\mu \in \mathfrak{O}_2^+$ *as*

$$\mu = (3^{\frac{1}{2}})^a \prod q_i^{b_i} \pi_j^{c_j} \rho_k^{d_k} \cdot \epsilon^e \tag{34.5}$$

(*where* q_i, π_j, *and* ρ_k *are of the types listed earlier in this section, and, of course, two different* π_i *or* ρ_k *can be conjugates*). *Then*

$$L(\mu) = L(3^{a/2}) \prod L(q_i^{b_i}) L(\pi_j^{c_j}) L(\rho_k^{d_k}) \tag{34.6}$$

if a *and all* d_k *are even, while*

$$L(\mu) = 0 \tag{34.7}$$

if a *or some* d_k *is odd.*

The primary components can each be treated by

$$L(\rho^d) = (-r)^{d/2} \textit{ and } L(3^{a/2}) = (-3)^{a/2} \tag{34.8}$$

for a *or* d *even, while starting with* $L(q)$ *or* $L(\pi)$.

$$\begin{aligned} L(q^b) &= L(q)L(q^{b-1}) - q^2 L(q^{b-2}) \\ L(\pi^c) &= L(\pi)L(\pi^{c-1}) - pL(\pi^{c-2}). \end{aligned} \tag{34.9}$$

35. Some empirical conjectures. We append a table of values of $L(q)$ and $L(\pi)$. The table was calculated by hand by Mr. T. Haskell and Mr. R. DeVore, using formula (28.9). The values of $L(2q)$ and $L(2\pi)$ were also calculated for checking by (34.4).

On the basis of these values we would venture the conjectures that

$$L(q) = 2q \text{ if } q \equiv 5 \pmod{12}, \tag{35.1}$$

$$L(\pi) \equiv \begin{cases} 10 \text{ or } 2 \text{ if } p \equiv 1 \pmod{24} \\ 14 \text{ or } 22 \text{ is } p \equiv 13 \pmod{24}. \end{cases} \tag{35.2}$$

In the case of $q \equiv -5 \pmod{12}$ the numerical evidence is very inadequate, but with greater temerity we observe that $L(q) \equiv 2 \pmod{24}$.

If this calculation were programmed for a computer, undoubtedly enough data could be gathered to make congruence conjectures which could be translated into identities on modular forms, as has been done by M. Newman for the partition function using the modular group in one complex variable.

Table of $L(q)$

$q \equiv 5 \pmod{12}$	L	$q \equiv -5 \pmod{12}$	L
5	10	7	2
17	34	19	26
29	58	31	-46
41	82	43	-22
53	106	67	122
...	...	79	-142

Table of $L(a + 2b3^{\frac{1}{2}})$ for $p = a^2 - 12b^2$

p	a	b	L	p	a	b	L
13	5	1	-2	421	23	3	22
37	7	2	-10	433	25	4	2
61	13	6	14	457	35	8	10
73	11	4	10	541	29	5	46
97	17	8	-14	577	25	2	-46
109	11	1	-2	601	37	8	26
157	13	1	14	613	25	1	-10
181	17	3	-26	661	31	5	38
193	25	6	2	673	49	12	50

p	a	b	L	p	a	b	L
229	23	5	22	709	41	9	22
241	17	2	—14	733	29	3	—50
277	17	1	—26	757	47	11	—26
313	19	2	—22	769	31	4	2
337	23	4	34	829	29	1	46
349	19	1	14	853	31	3	—58
373	31	7	38	877	43	9	—34
397	37	9	—34	937	37	6	26
409	29	6	—38	997	55	13	—10

36. Unit multiplier. We shall be concerned with the following series (summed over $\mu \in \mathfrak{O}_2^+$):

$$\Lambda(\tau) = \Lambda(0,0;\tau) = \sum_{\mu \in \mathfrak{O}_2^+} L(\mu) e(\mu\tau). \tag{36.1}$$

We first note, in accordance with equations (31.6), that $\Lambda(1,1;\epsilon\tau)$ as a function of τ belongs to $\mathfrak{H}^*$. It is symmetric, since it must equal $\Lambda(1,1;\epsilon'\tau)$, hence it is proportional to $\Lambda(0,0;\tau)$. Then by checking the leading coefficients (which requires a return to $\Theta_{1,1}^4(\tau)$ and $\Omega_{1,1}(\tau)$ of [16, § 6]), we see

$$\Lambda(1,1;\tau) = \Lambda(0,0;\epsilon\tau). \tag{36.2}$$

Next we consider the sum of the functions $\Lambda(0,0;\tau) + \Lambda(0,1;\tau) + \Lambda(1,0;\tau) + \Lambda(1,1;\tau)$. It is a cusp-form defined on $\mathfrak{G}$ hence on $\mathfrak{H}^*$; but since the non-vanishing cusp-form $\Lambda(\tau)$ is defined not on $\mathfrak{G}$ but on $\mathfrak{H}^*$, it follows that the sum is zero, or

$$0 = \Lambda(0,0;\tau) + \Lambda(0,1;\tau) + \Lambda(1,0;\tau) + \Lambda(1,1;\tau). \tag{36.3}$$

This last result can be simplified if we notice from (36.2) and (31.6)

$$\Lambda(0,1;\tau) = \Lambda(1,1;\tau+\eta) = \Lambda(0,0;\epsilon\tau+\epsilon\eta) + \Lambda(\epsilon\tau+\eta) \tag{36.4}$$

since $\epsilon\eta \equiv \eta \pmod 2$. Likewise

$$\Lambda(1,0;\tau) = \Lambda(\tau+\eta) \tag{36.5}$$

hence the identity (36.3) becomes

$$\Lambda(\tau) + \Lambda(\tau\epsilon) + \Lambda(\tau\epsilon+\eta) + \Lambda(\tau+\eta) = 0 \tag{36.6}$$

or, coefficientwise, summing over $\mu \in \mathfrak{O}_2^+$, we have

$$\begin{aligned} \sum L(\mu)e(\mu\tau) + L(\mu)e(\mu\epsilon\tau) &+ L(\mu)e(\mu\epsilon\tau)e(\mu\eta) \\ &+ L(\mu)e(\mu\tau)e(\mu\eta) = 0. \end{aligned} \tag{36.7}$$

Since $e(\mu\eta) = 1$ if $2 \mid \mu$ and -1 if $2 \nmid \mu$, (36.7) reduces on the left to $2 \sum_{\mu \in 2\mathfrak{D}^+} [L(\mu) + L(\mu\epsilon)] e(\mu\tau)$; hence

$$L(\mu) = -L(\mu\epsilon) \text{ if } \mu \in 2\mathfrak{D}^+. \tag{36.8}$$

37. Even multipliers. The most difficult identity is actually (34.4) involving multipliers of powers of η. We consider the "four-valued" functions

$$\begin{cases} P(\tau) = \Lambda(0,0;\tau) + \Lambda(1,1;\tau) - \Lambda(1,0;\tau) - \Lambda(0,1;\tau) \\ M(\tau) = \Lambda(0,0;\tau) - \Lambda(1,1;\tau) + \Lambda(1,0;\tau) - \Lambda(0,1;\tau) \\ K(\tau) = \Lambda(0,0;\tau) - \Lambda(1,1;\tau) - \Lambda(1,0;\tau) + \Lambda(0,1;\tau). \end{cases} \tag{37.1}$$

By the transformation equations (31.6) for $\Lambda(d, c; \tau)$ we find

$$(37.2)\quad \begin{cases} M(\tau+1) = M(\tau) \\ K(\tau+1) = P(\tau) \\ P(\tau+1) = K(\tau) \end{cases} \qquad (37.3)\quad \begin{cases} M(\tau+\eta) = M(\tau) \\ K(\tau+\eta) = -K(\tau) \\ P(\tau+\eta) = -P(\tau) \end{cases}$$

$$(37.4)\quad \begin{cases} M(\epsilon\tau) = -M(\tau) \\ K(\epsilon\tau) = -K(\tau) \\ P(\epsilon\tau) = P(\tau) \end{cases} \qquad (37.5)\quad \begin{cases} M(-1/\tau) = K(\tau)N(\tau)^2 \\ K(-1/\tau) = M(\tau)N(\tau)^2 \\ P(-1/\tau) = P(\tau)N(\tau)^2. \end{cases}$$

Note that by virtue of (36.2) the substitution $\tau_1 = \epsilon\tau$ can be evaluated thus making the *whole* group Ω accessible.

The important result, immediately, is that the identity (34.4) is contained coefficientwise precisely in $G(\tau) = 0$ where

$$\begin{aligned} G(\tau) = -\Lambda(\tau) - \Lambda(\tau+\eta) + 2\{\Lambda(2\tau) - \Lambda(2\epsilon\tau) \\ - \Lambda(2\tau+\eta) + \Lambda(2\epsilon\tau+\eta)\}. \end{aligned} \tag{37.6}$$

Actually $G(\tau)$ does not have enough useful invariance properties but we can set

$$H(\tau) = G(\epsilon\tau) - G(\tau) = M(\tau) - 4K(2\tau). \tag{37.7}$$

We shall eventually show $H(\tau) \equiv 0$ and we might pause just to notice that, with some minor rearranging,

$$\begin{aligned} H(\tau) = 4 \sum_{\nu \in \mathfrak{D}^+} L(2\nu) e(2\nu\tau) - 8 \sum_{\substack{\nu \in \mathfrak{D}_2^+ \\ \eta \nmid \nu}} L(\nu) e(2\nu\tau) \\ + 8 \sum_{\eta \nmid \nu \in \mathfrak{D}_2^+} L(\nu) e(2\nu\epsilon\tau). \end{aligned} \tag{37.8}$$

The reader can verify that the identity (34.4) follows from $H(\tau) \equiv 0$. We shall, however, have to go sightly further and form a function also invariant under $\tau_1 = -1/\tau$, namely

$$\Phi(\tau) = 4[H(\tau) + H(-1/\tau)/N(\tau)^2]. \tag{37.9}$$

Using the identities (37.5) we write

$$H(-1/\tau)/N(\tau)^2 = -H(\tau/2)/4. \tag{37.10}$$

Thus

$$\Phi(\tau) = 4H(\tau) - H(\tau/2), \tag{37.11}$$

and

$$\Phi(\tau) = 4M(\tau) - 16K(\tau) - M(\tau/2) + 4K(\tau). \tag{37.12}$$

Note that if $\Phi(\tau) \equiv 0$ then by (37.11) it would be necessary that $H(\tau) \equiv 0$, (since otherwise no order at ∞ could be defined for H in terms of Fourier series).

38. The vanishing of Φ. As a first step we verify that Φ is defined on $\mathfrak{J}$. We remark that by (37.12) and (37.2)-(37.5), $\Phi(\tau)$ is invariant under the last four matrices in (29.5). Furthermore $\Phi(\tau)$ is invariant under $\mathfrak{G}_4$. To see this we first note that each $\Lambda(d, c; \tau)$ is invariant under those elements of $\mathfrak{H}$ which are in $\mathfrak{G}_2$, namely the set of matrices congruent to $\mathfrak{e} \pmod 2$, (indeed the elements of $\mathfrak{G}_2$ leave the quantities (29.3) all unchanged). Hence $M(\tau)$, $K(\tau)$, and $P(\tau)$ are invariant under these matrices. Furthermore, if τ is transformed according to an element of $\mathfrak{G}_4$ then both 2τ and $\tau/2$ are transformed according to an element of $\mathfrak{G}_2$. We see this, for example, in the case of 2τ, if we write $\tau_1 = [\alpha\tau + \beta]/[\gamma\tau + \delta]$ as $(2\tau_1) = [\alpha(2\tau) + 2\beta]/[(2\tau)\gamma/2 + \delta]$.

We now use Theorem 29.1(d) to express

$$\mathfrak{H} = \mathfrak{J}\{\mathfrak{t}_1 + \mathfrak{t}_2 + \cdots + \mathfrak{t}_6\} \tag{38.1}$$

where

$$\begin{gathered}\mathfrak{t}_1 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad \mathfrak{t}_2 = \begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix} \quad \mathfrak{t}_3 = \begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}\\ \mathfrak{t}_4 = \begin{pmatrix}1+2\eta & 2\eta\\ -2\eta & 1-2\eta\end{pmatrix} \quad \mathfrak{t}_5 = \begin{pmatrix}1+2\eta & -1\\ -2\eta & 1\end{pmatrix} \quad \mathfrak{t}_6 = \begin{pmatrix}1 & 2\eta\\ -1 & 1-2\eta\end{pmatrix}\end{gathered} \tag{38.2}$$

We then form a new function,

$$F(\tau) = F(0, 0; \tau) = \prod_{k=1}^{6} \Phi(\mathfrak{t}_k\tau)/N(\operatorname{den} \mathfrak{t}_k\tau)^2. \tag{38.3}$$

This function is defined on $\mathfrak{H}$ and is of dimension -12, $(w = 6)$. It is a cusp-form by virtue of the known behavior of $\Lambda(d, c; \tau)$. We shall show that in its expansion (32.7) $A_\mu = 0$ when $a \leq 24$. By Theorem 32.2, this will complete the proof that $F(\tau) \equiv 0$ hence $\Phi(\tau) \equiv 0$, (since all $\mathfrak{t}_k\tau$ have the same range).

To form $F(1,1;\tau_1)$ from $F(0,0;\tau)$, we must set $\tau = [\tau_1 + \eta]/[\tau_1 + \epsilon]$ according to (31.4). But by the transformations permissible on Φ,

$$\Phi(\mathfrak{t}_2\tau) = \Phi(\tau - 1) = \Phi(-1/[\tau_1 + \epsilon]) = \Phi(\tau_1 + \epsilon) N(\tau_1 + \epsilon)^2 \tag{38.4}$$

and

$$\begin{aligned} \Phi(\mathfrak{t}_3\tau)/N(\operatorname{den}\mathfrak{t}_3\tau)^2 &= \Phi(\tau/[-\tau + 1])/N(-\tau + 1)^2 \\ &= \Phi(\tau_1 + \eta) N(\tau_1 + \epsilon)^2. \end{aligned} \tag{38.5}$$

Thus, re-grouping factors in (38.3), we isolate $\Phi(\tau_1 + \epsilon)$ and $\Phi(\tau_1 + \eta)$ yielding

$$F(1,1;\tau_1) = \Phi(\tau_1 + \epsilon)\Phi(\tau_1 + \eta)\Phi^*(\tau_1), \tag{38.6}$$

where Φ^* is a collection of the remaining factors. Dropping the subscript, we note (38.6) represents $F(1,1;\tau)$ as the product of three cusp-forms in τ, the first two of which are very familiar through the expansions (37.11) and (37.8).

We now verify by actual calculation that the identities (34.4) are valid when $\mu = a + 2b3^{\frac{1}{2}}$ satisfies the inequality $a \leqq 24$. Then in the Fourier expansion for $\Phi(\tau) = \sum A_\nu^* e(\nu\tau)$, $A_\nu^* = 0$ when $\nu = r + s3^{\frac{1}{2}}$ for $r \leqq 12$, and thus the expansion $F(1,1;\tau) = \sum A_\nu e(\nu\tau)$ has as non-zero terms only those A_ν where $\nu = r + s \cdot 3^{\frac{1}{2}}$ and $r \geqq 24$. Then from Theorem 32.2, $\Phi \equiv 0$. This calculation was quite lengthy and was therefore omitted. We included the table in § 35 (above), which may be a useful by-product.

39. Odd prime multipliers in $\mathfrak{O}_2^+$. We let κ denote a general prime of type q or π (in § 33); thus $\kappa \in \mathfrak{O}_2^+$. In accordance with the Mordell-Hecke method, we form

$$T_\kappa[\Lambda(\tau)] = \Lambda(\kappa\tau) + \sum_{i=1}^{|N(\kappa)|} \Lambda([\tau + \beta_i]/\kappa)/N(\kappa)^2 \tag{39.1}$$

where β_i runs over all $|N(\kappa)|$ residue classes modulo κ, chosen, however, so that $\beta_i \in \mathfrak{O}_2$, (which can be done since κ is odd). Subject to this condition, $T_\kappa[\Lambda(\tau)]$ is independent of the choice of β_i, in fact if $\beta - \beta^* \in \kappa\mathfrak{O}_2$, then $\Lambda([\tau + \beta]/\kappa) = \Lambda([\tau + \beta^*]/\kappa)$ by the fact that $\Lambda(\tau)$ belongs to $\mathfrak{H}$.

Then it is immediate that $T_\kappa[\Lambda(\tau)]$ is invariant under the substitution $\tau_1 = \tau + 1$, $\tau_1 = \tau + 2\eta$, and $\tau_1 = \epsilon^2\tau$. To see the effect of $\tau_1 = -1/\tau$, note $\Lambda(-\kappa/\tau) = \Lambda(\tau/\kappa)N(\tau/\kappa)^2$, while $\Lambda([\tau_1 + \beta]/\kappa) = \Lambda([-1 + \beta\tau_1]/\kappa\tau)$. But if β is one of the above residue classes prime to κ, then we can find a ξ_1, ξ_2, ξ_3, ξ_4 in $\mathfrak{O}_2$ such that

$$[\xi_1(-1+\beta\tau)+\xi_2\kappa\tau]/[\xi_3(-1+\beta\tau)+\xi_4\kappa\tau]=[\tau+\beta^*]/\kappa \tag{39.2}$$

where $\xi_1\xi_4-\xi_2\xi_3=1$ and β^* is defined by $-\beta\beta^*\equiv 1 \bmod \kappa$, $\beta^*\in\mathfrak{O}$. Indeed if $-\beta\beta^*+\xi\kappa=1$ we can let $\xi_1=-\beta^*$, $\xi_2=\xi$, $\xi_3=-\kappa$, $\xi_4=\beta$.

We can perform an explicit series expansion of $T_\kappa[\Lambda(\tau)]$ using (36.1) and the "polygon" theorem [25, § 18]:

$$\sum_{\substack{\beta\in\mathfrak{O}_2\\ \bmod \kappa}} e(\mu\beta/\kappa)=\begin{Bmatrix} 0 \text{ if } \kappa\nmid\mu \\ |N(\kappa)| \text{ if } \kappa\mid\mu\end{Bmatrix}=I(\mu/\kappa)\,|N(\kappa)|, \tag{39.3}$$

where $I(\cdot\cdot\cdot)$ is 1 or 0 depending on whether or not its argument is an integer. Thus

$$T_\kappa[\Lambda(\tau)]=\sum_{\mu\in\mathfrak{O}_2^+}\{L(\mu)e(\kappa\mu\tau)+L(\mu\kappa)e(\mu\tau)/|N(\kappa)|\} \tag{39.4}$$

Hence $T_\kappa[\Lambda(\tau)]$ is symmetric and proportional to $\Lambda(\tau)$. The factor of proportionality is immediate if we examine the term in $e(\tau)$, in fact,

$$T_\kappa[\Lambda(\tau)]=[L(\kappa)/|N(\kappa)|]\Lambda(\tau). \tag{39.5}$$

We henceforth define $L(\mu)=0$ if μ is any element of the field of $3^{\frac{1}{2}}$ which is not in $\mathfrak{O}_2^+$. Then the identities (39.4) and (39.5) can be written (for $\mu\in\mathfrak{O}_2^+$),

$$L(\mu/\kappa)\,|N(\kappa)|+L(\mu\kappa)=L(\mu)L(\kappa). \tag{39.6}$$

Following the method of Mordell [31], we can take $\mu=\kappa^t$ and inductively establish, (for $t\geqq 1$ and $\sigma\in\mathfrak{O}_2^+$),

$$L(\kappa^{t-2})\,|N(\kappa)|+L(\kappa^t)=L(\kappa)L(\kappa^{t-1}), \tag{39.7}$$

and if σ is prime to κ,

$$L(\kappa^{t-2}\sigma)\,|N(\kappa)|+L(\kappa^t\sigma)=\mathrm{L}(\kappa)L(\kappa^{t-1}\sigma). \tag{39.8}$$

Thus $L(\kappa^t\sigma)=L(\kappa^t)L(\sigma)$, which is part of the multiplication theorem.

40. Modification of Hecke's Operators. It is clear that if κ and λ are two primes of type q or π (not necessarily distinct) then the operators T_κ and T_λ must commute, (as would happen in the one-variable theory). Indeed

$$T_\kappa[T_\lambda[\Lambda(\tau)]]=[L(\kappa)L(\lambda)/|N(\kappa)N(\kappa)|]\Lambda(\tau). \tag{40.1}$$

The explicit formula is of some later interest:

$$\begin{aligned} T_\kappa[T_\lambda[\Lambda(\tau)]] &= \Lambda(\kappa\lambda\tau)+\sum_\beta \Lambda([\lambda\tau+\beta]/\kappa)/N(\kappa)^2 \\ &+\sum_\lambda \Lambda([\kappa\tau+\kappa\gamma]/\lambda)/N(\lambda)^2+\sum_{\beta,\lambda}\Lambda([(\tau+\gamma)/(\lambda+\beta)]/\kappa)/N(\kappa)^2N(\lambda)^2, \end{aligned} \tag{40.2}$$

where β and γ are residues $(\bmod\,\kappa)$ and $(\bmod\,\lambda)$ respectively and are chosen in $\mathfrak{O}_2$. Summing by coefficients, as in the last section, we obtain the following identity (making frequent use of (39.3)):

$$\begin{aligned}(40.3)\qquad \sum_{\mu\in\mathfrak{O}_2^+}\{L(\mu)e(\kappa\lambda\mu\tau)&+L(\mu)e(\lambda\mu\tau/\kappa)I(\mu/\kappa)/|N(\kappa)|\\ &+L(\mu)e(\tau\mu\kappa/\lambda)I(\kappa\mu/\lambda)/|N(\lambda)|\\ &+L(\mu)e(\tau\mu/\kappa\lambda)I(\mu/\kappa\lambda)/|N(\kappa)N(\lambda)|\}\\ &=\sum_{\mu\in\mathfrak{O}_2^+}L(\kappa)L(\lambda)L(\mu)e(\mu\tau)/|N(\kappa)N(\lambda)|.\end{aligned}$$

We then obtain the following analogue of (39.6):

$$\begin{aligned}(40.4)\qquad L(\mu/\kappa\lambda)&+L(\kappa\mu/\lambda)I(\mu/\lambda)/|N(\kappa)|+L(\lambda\mu/\kappa)I(\lambda\mu/\kappa)/|N(\lambda)|\\ &+L(\kappa\lambda\mu)/|N(\kappa)N(\lambda)|=L(\kappa)L(\lambda)L(\mu)/|N(\kappa)N(\lambda)|,\end{aligned}$$

for μ in $\mathfrak{O}_2^+$. The apparent "lack of symmetry" in the $I(\cdot\ \cdot\ \cdot)$ is needed to take care of the case where $\lambda=\kappa$.

Let us now change our assumption slightly, and let κ and λ be odd primes for which

$$(40.5)\qquad \kappa>0>\kappa',\qquad \kappa\in\mathfrak{O}_2,$$

and likewise for λ, whence $\kappa\lambda\in\mathfrak{O}_2^+$. For this purpose we would choose κ and λ to be primes of type ρ or $3^{\frac{1}{2}}\epsilon$. Clearly $T_\kappa[\Lambda(\tau)]$ can not be defined as in § 39, because the imaginary parts of τ and τ' would not maintain sign. *We assert, nevertheless, that the identities* (40.3) and (40.4), *obtained by joining* (40.1) *and* (40.2) *are still valid,* (as though the composition of two "invalid" Hecke operators were "valid"). To see this, note that in the case where κ and λ are in $\mathfrak{O}_2^+$ the validity of (40.1) and (40.2) is due to an explicit use of linear transformations of type (39.2). The existence of these transformations did not depend in any way on whether or not κ or λ lay in $\mathfrak{O}_2^+$ as long as they were in $\mathfrak{O}_2$. The only difficulty was that the imaginary parts of the arguments in (39.1) had the wrong signs. The signs, however, come out right in (40.2) as long as $\kappa\lambda\in\mathfrak{O}^+$. We therefore know that $T_\kappa[T_\lambda[\Lambda(\tau)]]$ is equal to $\Lambda(\tau)$ times a constant, (say) $C(\kappa,\lambda)$.

If we evaluate $C(\kappa,\lambda)$ by examining the term in $e(\tau)$ on the left hand side of (40.3) we see

$$(40.6)\qquad |N(\kappa\lambda)|\,C(\kappa,\lambda)=\begin{cases}L(\kappa\lambda) & \text{if } \kappa\neq\lambda\\ L(\kappa^2)+|N(\kappa)| & \text{if } \kappa=\lambda.\end{cases}$$

Actually we shall see in the next section that $C(\kappa,\lambda)=0$ whenever κ and λ satisfy (40.5). Thus, according to our extended definition, $L(\kappa)=L(\lambda)=0$ justifying (40.1) etc., in the literal sense.

We therefore regard (40.4) as a formula for relating $L(\mu)$ to $L(\mu\kappa\lambda)$ or for adjoining two prime factors at once. The basic analogue to (39.8) becomes (with μ prime to $\kappa\lambda$)

$$\begin{aligned} L(\kappa^{2t+1}\lambda^{3u+1}\mu) &+ |N(\lambda)|\, L(\kappa^{2t+1}\lambda^{2u-1}\mu) \\ &+ |N(\kappa)|\, L(\lambda^{2u+1}\kappa^{2t-1}\mu) + |N(\kappa)N(\lambda)|\, L(\kappa^{2t-1}\lambda^{2u-1}\mu) = 0 \end{aligned} \tag{40.7}$$

From this, the multiplicative property is immediate, as in the last section, while (34.8) can be built up similarly, by induction, starting with (40.6) with $C(\kappa,\lambda) = 0$. Most of the details must be left to the reader.

41. Evaluating the constant.

We now consider

$$T_\kappa[T_\lambda[\Lambda(\tau)]] = C(\kappa,\lambda)\Lambda(\tau) \tag{41.1}$$

with κ and λ odd primes subject to the conditions of (40.5). We wish to show $C(\kappa,\lambda) = 0$, and in any specific case we can verify this value by calculating the coefficient of $e(\tau)$ on the left-hand side of (40.3). We note, however, that

$$\begin{cases} C(\kappa_1,\lambda_1)C(\kappa_2,\lambda_2) = C(\kappa_1,\lambda_2)C(\kappa_2,\lambda_1) \\ C(\kappa,\lambda) = C(\lambda,\kappa). \end{cases} \tag{41.2}$$

These *associativity* and *commutativity* properties of Hecke's operators can first be established by inspection of (40.1) and (40.2) for κ_1, λ_1, κ_2, λ_2 in $\mathfrak{O}_2^+$. Then, all we need do is note that the same linear transformations are still present when κ_1, λ_1, κ_2, λ_2 are not in $\mathfrak{O}_2^+$ but are taken more generally.

Consequently, if $C(\kappa,\kappa) = 0$, then $C(\kappa,\lambda) = 0$ since $C(\kappa,\lambda)C(\kappa,\lambda) = C(\kappa,\kappa)C(\lambda,\lambda)$. We can verify directly that $C(3^{\frac{1}{2}}\epsilon, 3^{\frac{1}{2}}\epsilon) = 0$ hence, from now on, κ and λ are chosen to be strictly of type ρ (see § 33). We now need only show $C(\kappa,\kappa) = 0$. Since $C(\kappa,\kappa) = C(\kappa',\kappa')$ then $C(\kappa,\kappa)C(\kappa',\kappa') = C(\kappa,\kappa')^2$ and it suffices to show $C(\kappa,\kappa') = 0$, or, by equation (40.6), it suffices to show $L(\kappa\kappa') = 0$. If we let $k = |N(\kappa)|$, this amounts to the statement, (from (28.9)), that if k is a prime $\equiv -1 \pmod{12}$,

$$A_4(k) = 4D^*(k) = 4(1+k)^2 \tag{41.3}$$

Actually using a method of Pall, this last relation can be proved directly by transforming equation (27.1), (with $\nu = k$), into the rational equation

$$k^2 = t_0^2 + 12(t^2 + t_2^2 + t_3^2) \tag{41.4}$$

and showing this equation to have $(1+k^2)/2$ solutions. Here, although the quarternary form in (41.4) belongs to a genus with two classes, Pall has shown *the number of representations by each class is the same* under certain conditions arising from $k=\kappa\kappa'$ in (27.1), effectively simulating the condition of the single class in the genus. We do not give details here because of joint work to appear soon [26]. It would be desirable to prove $C(\kappa,\lambda)=0$ completely by modular functions, but no such method seems apparent at present.

42. The zero manifold. We conclude with some remarks on the function $\Lambda(\tau,\tau')=\Lambda(0,0;\tau,\tau')$ on the zero manifold $\tau=-\tau'=\mu$, as described in § 34 (above). We call

$$\Lambda^{(t)}(u)=(\partial/\partial v)^t\Lambda(1,1;\tau) \tag{42.1}$$

and from $\Lambda(1,1;t)=\Lambda(\epsilon\tau)$ it follows that

$$\Lambda^{(t)}(u)=(\pi i)^t\sum_{a=1}^{\infty}\exp \pi i a/3^{\frac{1}{2}}\cdot\sum_{(b)}L[(a+b3^{\frac{1}{2}})(2+3^{\frac{1}{2}})]b^k, \tag{42.2}$$

summed over all b for which $a+b3^{\frac{1}{2}}$ is totally positive, and subject to the convention $0^0=1$.

Now $\Lambda^{(0)}(u)=0$ by [16, § 10] leading to the identity

$$L(2a)=2\sum_{b=1}^{[2a/3^{\frac{1}{2}}]}L[(2a+b3^{\frac{1}{2}})(2+3^{\frac{1}{2}})]. \tag{42.3}$$

On the other hand, we can write

$$\Lambda^{(2)}(u)=(\pi i)^2\sum_{1}^{\infty}\xi(a)\exp 2\pi i a u/3^{\frac{1}{2}}. \tag{42.4}$$

where

$$\xi(a)=\sum_{b=1}^{2a/3^{\frac{1}{2}}}L[(2a+b3^{\frac{1}{2}})(2+3^{\frac{1}{2}})]b^2. \tag{42.5}$$

Here $\xi(1)=1$, $\xi(2)=6$, $\xi(3)=-27$, $\xi(5)=498$, etc. We can verify, again, by Mordell's method [31] that $\xi(a)$ is multiplicative. The basic equation is

$$p^7\xi(a/p)+\xi(pa)=\xi(p)\xi(a) \tag{42.6}$$

if p is a prime $\neq 3$ and the definition of ξ is extended to zero for non-integral arguments, while even if $3\mid a$

$$\xi(3a)=\xi(3)\xi(a). \tag{42.7}$$

From these equations, by the same method, we see

$$\xi(3^n) = \xi(3)^n \tag{42.8}$$

and

$$\xi(p^{n+1}) = \xi(p^n)\xi(p) - p^7\xi(p^{n-1}),\ p \neq 3; \tag{42.9}$$

and, finally, if $(a, b) = 1$, then

$$\xi(ab) = \xi(a)\xi(b). \tag{42.10}$$

To see equation (42.6), we construct the operator

$$U_p\phi(u) = \phi(pu) + \sum_{b=0}^{p-1} \phi([u + b3^{\frac{1}{2}}]/p)/p^8 \tag{42.11}$$

for p a prime $\neq 3$. Then we show that if $\phi(u)$ transforms like

$$\phi(u + 3^{\frac{1}{2}}) = \phi(u) \tag{42.12}$$

and

$$\phi(-1/u) = \phi(u)u^8 \tag{42.13}$$

then so does $U_p\phi(u)$. The non-trivial step consists in showing that if $0 < k < p$, and $-3kl \equiv 1 \pmod p$, then, with $f(u) = \phi([u + k \cdot 3^{\frac{1}{2}}]/p)$, $f(-1/u) = \phi([u + l \cdot 3^{\frac{1}{2}}]/p) \cdot (pu)^8$. In fact from $-3kl = 1 - Bp$ we can write

$$\frac{-l \cdot 3^{\frac{1}{2}}(-1 + ku \cdot 3^{\frac{1}{2}})/pu + B}{-p(-1 + ku \cdot 3^{\frac{1}{2}})pu + k \cdot 3^{\frac{1}{2}}} = \frac{u + l \cdot 3^{\frac{1}{2}}}{p} \tag{42.14}$$

while the transformation $u = (A \cdot 3^{\frac{1}{2}}u + B)/(Cu + D3^{\frac{1}{2}})$ can be shown (see [30]) to be generated by the transformations shown in (42.12) and (42.13). Now we saw [16, § 10] all cusp solutions to (42.12) and (42.13) are proportional to a single one, say $\Lambda^{(2)}(u)$, hence, finding the proportionality constant from the term in $\exp 2\pi iu/3^{\frac{1}{2}}$, we see

$$U_p\Lambda^{(2)}(u) = \xi(p)\Lambda^{(2)}(u)/p^7. \tag{42.15}$$

The relation (42.6) comes from the Fourier series.

To take care of $p = 3$, as in (42.7), we must examine the individual terms of $\xi(a)$, (although, a proof using only Mordell-Hecke operations would still be of some interest). From (42.5), if $3 \nmid a$,

$$\xi(3^t a) = \sum_{b=1}^{\infty} L[(2a \cdot 3^t + b \cdot 3^{\frac{1}{2}})(2 + 3^{\frac{1}{2}})]b^2, \tag{42.16}$$

(using the extended definition of L for convenience). Let $b = 3^s r$, and

$$\gamma = (2a \cdot 3^t + 3^s r \cdot 3^{\frac{1}{2}}) = \begin{cases} 3^s \cdot 3^{\frac{1}{2}}\delta_1 & \text{if } t > s \\ 3^t \cdot \delta_2 & \text{if } t \leqq s, \end{cases} \tag{42.17}$$

where $3^{\frac{1}{2}} \nmid \delta_1$ or δ_2. By Theorem 34.1, $L(\gamma\epsilon) = 0$ if $t > s$, hence $3^t \mid b$ and if we set $b = 3^t B$, $\delta_2 = (2a + 3^{\frac{1}{2}}B)$, and $L(\gamma\epsilon) = L(3^t)L([2a + 3^{\frac{1}{2}}B](2 + 3^{\frac{1}{2}}))$ whence,

$$\xi(3^t a) = (-27)^t \xi(a)\,; \tag{42.17}$$

completing the multiplicativity.

Here, Hecke's Euler-product form [29] can be applied in the conventional manner:

$$\sum_1^\infty \xi(a) a^{-s} = (1 + 27 \cdot 3^{-s})^{-1} \prod_{p \neq 3} (1 - \xi(p) p^{-s} + p^{7-2s}). \tag{42.18}$$

43. Concluding remarks. In brief, the preceding investigation of the cusp function $\Lambda(u)$ suggests three major lines of development.

First of all the complex multiplication structure in §§ 36, 37 suggests a possibility of extending the class-field theory to totally complex quadratic extensions of $R(3^{\frac{1}{2}})$ by singular moduli in an explicit manner as was done in the quadratic case [27] and outlined by Hecke for the biquadratic case [17].

Secondly this same structure makes available the functions $\psi_1(\tau, \tau') = M(\tau) : K(\tau)$ and $\psi_2(\tau, \tau') = K(\tau) : P(\tau)$ which are invariants (to within a finite number of values) under $\mathfrak{O}$. Actually, we can just use

$$K(\tau) = 2 \sum_{\eta \nmid \nu \in \mathfrak{O}_2^+} L(\nu)[e(\nu\tau) - e(\nu\epsilon\tau)] \tag{43.1}$$

since, by (37.2) and (37.7),

$$P(\tau) = K(\tau + 1), \quad M(\tau) = 4K(2\tau). \tag{43.2}$$

We can also obtain absolute invariants Ψ_1, Ψ_2 under $\mathfrak{O}$ and in this way obtain a pseudo-conformal map of the fundamental domain of $\mathfrak{O}$ in τ, τ' space onto a finite covering of the Ψ_1, Ψ_2 space in various ways. This opens the possibility that we might define invariants by pseudo-conformal mapping as done in one dimension for the Klein [2] and Hecke [28] invariants as in Poincaré [33]). At any rate the topology of the covering could reveal the dimension of the modular forms. The use of the Bergman kernel function [24] by A. Selberg [34] is indicative of the further possibilities of geometric function theory.

Finally, the vital role of the zero-manifold is perhaps related to the projection of the representation by sums of squares in $R(3^{\frac{1}{2}})$ into rational quaternary problems which Pall has shown how to use to advantage [26].

Added in proof: Dr. H. Braun has kindly called attention to work of Dzewas [35] consummating Siegel's genus method (see § 27 above) but not overlapping the methods of the present series of papers.

Errata for Part I (i. e., [16]).

Page 301 equ. (1),	$\lvert b\rvert^{\frac{1}{2}}$	should read	$2\lvert b\rvert^{\frac{1}{2}}$
" 312 equ. (47c),	$\psi_0(\tau)$	" "	$\psi_0(2\tau)$
" 316 equ. (62b),	$\Gamma_j(u)$	" "	$\Gamma_{j-1}(u)$
	$\Gamma_{j+1}(u)$	" "	$\Gamma_j(u)$
" 319 equ. (78b)	$\Gamma_j(u)$	" "	$\Gamma_{j-1}(u)$
	$\Gamma_{j+1}(u)$	" "	$\Gamma_j(u)$
" 320 line 7	$\Xi_{1,1}(\tau)/\Theta_{1,1}(\tau)$	should read	$\Xi_{1,1}(\tau)/\Theta_{1,1}{}^4(\tau)$

Errata for Part II (i. e., [25]).

Page 44 line 26,	one (not both)	should read	both
" 49 equ. (22.8),	$\Gamma_j(u)$	" "	$\Gamma_{j-1}(u)$
	$\Gamma_{j+1}(u)$	" "	$\Gamma_j(u)$

UNIVERSITY OF ARIZONA.

REFERENCES.

[16] H. Cohn, "Decomposition into four integral squares in the fields of $2^{\frac{1}{2}}$ and $3^{\frac{1}{2}}$," *American Journal of Mathematics*, vol. 82 (1960), pp. 301-321.

[24] S. Bergman, *The Kernel function*, New York (1950).

[25] H. Cohn, "Calculation of Class numbers by decomposition into three integral squares in the fields of $2^{\frac{1}{2}}$ and $3^{\frac{1}{2}}$," *American Journal of Mathematics*, vol. 83 (1961), pp. 33-56.

[26] H. Cohn and G. Pall, "Sum of four squares in a quadratic ring," (in preparation).

[27] R. Fueter, *Vorlesungen über die singulären Moduln und die komplexe Multiplikation der elliptischen Funktionen*, vols. 1, 2, Leipzig-Berlin (1924).

[28] E. Hecke, "Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung," *Mathematische Annalen*, vol. 112 (1936), pp. 664-699.

[29] ———, "Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung, I, II," *Mathematische Annalen*, vol. 114 (1937), pp. 1-28, pp. 316-351.

[30] J. I. Hutchinson, "On a class of automorphic functions," *Transactions of the American Mathematical Society*, vol. 3 (1902), pp. 1-11.

[31] L. J. Mordell, "On Mr. Ramanujan's Empirical Expansions of Modular Functions," *Proceedings of the Cambridge Philosophical Society*, vol. 19 (1917), pp. 117-124.

[32] G. Pall and O. Taussky, "Application of quaternions to the representation of a binary quadratic form as a sum of four squares," *Proceedings of the Royal Irish Academy*, section A, vol. 58 (1957), pp. 23-28.

[33] H. Poincaré, "Fonctions modulaires et fonctions Fuchsiennes," *Oeuvres*, vol. II, Paris (1916), pp. 592-618.

[34] A. Selberg, "Automorphic functions and integral operators," *Seminars on analytic functions, Institute for Advanced Study* 1957, Report TN 58-750 AD 201-502, U. S. Air Force Office of Scientific Research and Development Command, vol. II, pp. 152-159.

[35] J. Dzewas, "Quadratsummen in reell-quadratischen Zahlkörpern," *Mathematische Nachrichten*, vol. 21 (1960), pp. 233-284.

A GENERALIZATION OF A THEOREM OF HECKE.*[1]

By DONALD L. MCQUILLAN.

Introduction. Suppose K denotes the field of elliptic modular functions of prime level p, and j the Weierstrass absolute invariant. Then K is a Galois extension of $\boldsymbol{C}(j)$ with $LF(2,p)$ for Galois group and every element of this group gives rise to an automorphism of the vector space of differentials of the first kind in K. In this way we get a representation of $LF(2,p)$ with corresponding character X say. In two papers (cf. [3], [4]), Hecke determined this character completely, i. e. he determined how X decomposes into simple characters. When $p \equiv 3 \bmod 4$, $LF(2,p)$ has exactly two simple complex characters which happen to be complex conjugate, and Hecke discovered that the difference of the multiplicities of these characters in X yields the Dirichlet expression for $h(-p)$, the class-number of $\boldsymbol{Q}(\sqrt{-p})$. However, Hecke derived this remarkable fact by computation and the underlying reason for it is not yet known. With this problem in mind we shall investigate whether Hecke's result is restricted to prime levels p satisfying $p \equiv 3 \bmod 4$, or not. It turns out that there exists a rather natural generalization for those levels n such that $-n$ is a "fundamental discriminant." More precisely, if $n = \prod_{i=1}^{t} p_i$ where the p_i are distinct odd primes and $n \equiv 3 \bmod 4$, $n > 3$, then $LF(2,n)$ has 2^{2t-1} simple characters of class (G) (see definition in Section 1) and a certain alternating sum of the multiplicities of these characters in X equals $2^{t-1}h(-n)$.

We shall say a few words about our method. Let $K_0 \subset K$ be the fixed field of the cyclic group (T) generated by $T = \pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then, of course, K is a Galois extension of K_0 of degree n with (T) for Galois group. Using the Riemann-Roch theorem, (cf. [2], [8]), the multiplicities of the simple characters of (T) in $\operatorname{res}_{(T)}X$ are computed. This is done in Section 2. Then, denoting the simple characters of $LF(2,n)$ by $\{\chi\}$ and the multiplicity of χ in X by m_χ we have $X = \sum_\chi m_\chi \chi$ and consequently $\operatorname{res}_{(T)}X = \sum_\chi m_\chi \operatorname{res}_{(T)}\chi$. The relation between the m_χ and the multiplicities computed in Section 2

---

* Received November 16, 1961.

[1] This paper is part of the author's Ph. D. thesis written under the direction of Professor J.-I. Igusa at the Johns Hopkins University.

will then be found in Section 3 and the desired generalization of Hecke's result is obtained. The case where $-n$ is an even discriminant may be found in my dissertation. Also we note that the methods used are elementary and purely algebraic.

1. Preliminaries. Let τ be a complex variable with positive imaginary part and n a positive integer. Then the transformation $\tau \to (a\tau + b)/(c\tau + d)$, where a, b, c, d are integers, $ad - bc = 1$, $a \equiv d \equiv 1 \bmod n$, $b \equiv c \equiv 0 \bmod n$, is an analytic automorphism of the upper-half plane and the group of such automorphisms has a fundamental domain which can be compactified to a Riemann surface. A complex-valued function f in the upper-half plane, which is invariant under this "principal congruence group modulo n" and whose singularities are algebraic in the sense f becomes a meromorphic function on the compact Riemann surface is called a *modular function of level* n. For instance the Weierstrass invariant j is a modular function of level 1 and, furthermore, every modular function of level 1 is a rational function of j with complex coefficients. In general, the modular functions of level n form an algebraic function field of one variable whose field of constants is the field $\boldsymbol{C}$ of complex numbers. We denote this field by K. Also, we shall denote by $SL(2,n)$ the special linear group of degree two with coefficients in Z/nZ and define $LF(2,n)$ by

$$LF(2,n) = SL(2,n)/\pm\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

The order of $LF(2,n)$ is $\frac{1}{2}n\phi(n)\psi(n)$, where $\phi(n)$ is the Euler function and

$$\psi(n) = n\prod_{p|n}(1 + 1/p).$$

The cases $n = 1, 2$ are exceptional and the orders are then $n\phi(n)\psi(n) = 1, 6$. The following result is basic: *The field K of modular functions of level n is a finite Galois extension of $\boldsymbol{C}(j)$ with $LF(2,n)$ for Galois group and K is ramified over $\boldsymbol{C}(j)$ only at $j = 0$, 12^3, ∞ with ramification indices 3, 2, n respectively. Furthermore the inertia groups over these places are conjugate respectively to the cyclic groups generated by* $\pm\begin{pmatrix}-1 & -1\\ 1 & 0\end{pmatrix}$, $\pm\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$, $\pm\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ *in* $LF(2,n)$. Actually, this structure of K as an extension of $\boldsymbol{C}(j)$ is all we need in the following. We note also that these properties of the field of modular functions of level n are proved algebraically [7].

In the following we shall assume that $n = p_1p_2\cdots p_t$ where the p_i are distinct odd primes for $i = 1, 2, \cdots, t$, and that $n \equiv 3 \bmod 4$; $-n$ is then a fundamental discriminant which we shall denote by d. Using the Legendre symbol (a/p), we introduce the Kronecker symbol (d/a) as

$$(d/a) = \prod_{i=1}^{t} (a/p_i).$$

Definition. Let a be an integer which is relatively prime to n. If $(a/p_i) = e_i$ for $i = 1, 2, \cdots, t$, we shall say that a is of type $(e) = (e_1, e_2, \cdots, e_t)$, written $a \sim (e)$ and call $s(e)$ the number of $e_i = -1$ where $i = 1, 2, \cdots, t$. It is clear that if $a \sim (e)$ then $(d/a) = (-1)^{s(e)}$.

We note that the correspondence $a \rightarrow (e)$ gives rise to a homomorphism of the group of units of Z/nZ to an abelian group of type $(2, 2, \cdots, 2)$ (t-factors). By the Chinese remainder theorem, we see that this homomorphism is surjective and that $a \bmod n$ is in the kernel if and only if there exists a solution of the congruence $x^2 \equiv a \bmod n$. Since the group of units of Z/nZ is of order $\phi(n)$, the kernel consists of $\phi(n)/2^t$ classes. We shall denote representatives of these classes by $\{a_i\}$, where $i = 1, 2, \cdots, \phi(n)/2^t$.

Now the group $SL(2,n)$ is isomorphic to the direct product of the $SL(2,p_i)$, $i = 1, 2, \cdots, t$ under the map which sends $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \bmod n$ to $\prod_{i=1}^{t} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bmod p_i$. Therefore, by a general theorem, all simple characters χ of $SL(2,n)$ are of the form $\prod_{i=1}^{t} \chi_i$ where χ_i is a simple character of $SL(2,p_i)$ for $i = 1, 2, \cdots, t$, and conversely. A modified table of characters of $SL(2,p)$ appears at the end of the paper; it is adapted from a complete table as it appears in [12]. We note that the values of the characters at $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $-\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ differ at most by sign, and also that there are four characters with values which are quadratic irrationalities, i.e. $\psi_\pm$, $\psi_\pm'$. We shall call these the *characters of class* (G) of $SL(2,p)$. Let $E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in $SL(2,n)$, $E_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in $SL(2,p_i)$, $F = -E$, $F_i = -E_i$. The simple characters of $LF(2,n)$ are precisely those simple characters of $SL(2,n)$ for which $\chi(E) = \chi(F)$ and consequently $\prod_{i=1}^{t} \chi_i$ is a simple character of $LF(2,n)$ if and only if

$$\prod_{i=1}^{t} \chi_i(E_i) = \prod_{i=1}^{t} \chi_i(F_i).$$

A character $\prod_{i=1}^{t} \chi_i$ of $LF(2,n)$ will be called a *character of class* (G) if χ_i is a character of class (G) of $SL(2,p_i)$ for $i=1,2,\cdots,t$. The number of such characters is

$$2^t+\binom{t}{2}2^{t-2}2^2+\cdots\cdots=2^{2t-1}.$$

2. Generalized Hecke Formula (preliminary form). Let (T) be the cyclic subgroup of $LF(2,n)$ generated by $T=\pm\begin{pmatrix}1&1\\0&1\end{pmatrix}$ and K_0 the corresponding subfield of K. Let Ω be the vector space of differentials of the first kind in K. Then every automorphism of K gives rise to an automorphism of Ω and in this way we get representations of $LF(2,n)$ and (T). The corresponding characters will be called X and X_0 respectively. Of course $X_0=\operatorname{res}_{(T)}X$. Now (T) has n simple characters ψ_a defined by $\psi_a(T)=e(a/n)$, $a=0,1,\cdots,n-1$, where $e(x)=\exp(2\pi ix)$. In this section the multiplicity N_a of ψ_a in X_0 is determined when $(a,n)=1$ and the connection between these multiplicities and $h(-n)$ is established.

Some lemmas are needed. A typical element of $LF(2,n)$ will be denoted by $\pm\begin{pmatrix}a&b\\c&d\end{pmatrix}$.

Lemma 1. *Let r be a divisor of n. Then the number of elements of $LF(2,n)$ such that $(c,n)=r$ is $n^2\phi(n)/2r$.*

Proof. If $r=n$ the result is immediate. If $r=1$ then $b\equiv c^{-1}(ad-1)$ $\bmod n$ and the result is again immediate. Suppose now that $1<r<n$, and for clarity say $r=p_1p_2\cdots p_s$ with $1\leqq s<t$. The natural homomorphism from $LF(2,n)$ to $\prod_{i=1}^{t}LF(2,p_i)$ defined by $\pm\begin{pmatrix}a&b\\c&d\end{pmatrix}\bmod n\rightarrow\prod_{i=1}^{t}\pm\begin{pmatrix}a&b\\c&d\end{pmatrix}$ $\bmod p_i$ is surjective and has kernel

$$\{\pm\begin{pmatrix}a&0\\0&a\end{pmatrix}\mid a\equiv\pm1\bmod p_i\text{ for }i=1,2,\cdots,t\},$$

which is of order 2^{t-1}. Consider the image of those elements for which $(c,n)=r$. It consists of all those elements of the direct product in which the first s factors have $c\equiv0$ and the remaining factors have $c\not\equiv0$. The number of these elements is $n\phi(n)p_{s+1}p_{s+2}\cdots p_t/2^t$. Multiplying by the order of the kernel gives the desired result.

In order to state the next lemma, we introduce a notation. If x is a real number then $\langle x\rangle$ denotes the fractional part of x i.e. $\langle x\rangle=x-[x]$, $[x]$ being the largest integer $\leqq x$.

Lemma 2. *Let $d=-n$ and let r be any divisor of n different from n.*

Then $\sum(d/a)\langle a/r\rangle = 0$ *where* (d/a) *is the Kronecker symbol and the summation extends over a complete residue system* mod n *which are relatively prime to* n.

Proof. If $r = 1$, there is nothing to prove. Suppose therefore that $1 < r < n$. Without loss of generality we can take $r = p_1p_2\cdots p_s$ with $1 \leqq s < t$. Let $B = \pm r$ be a (fundamental) discriminant and $C = d/B$. A complete set $\{a\}$ of residues mod n which are relatively prime to n is obtained by taking $a = cB + bC$ with $0 < b < |B|$ and $(b,B) = 1$, $0 < c < |C|$ and $(c,C) = 1$. We then have $(d/a) = \epsilon(B/b)(C/c)$ where $\epsilon = (B/C)(C/B)$.

To obtain the members of $\{a\}$ for which $(d/a) = 1$ we can choose b arbitrarily but for a fixed b, c must be chosen so that $(C/c) = \epsilon(B/b)$. For a fixed b there are $\phi(|C|)/2$ such c. A similar statement holds for the members of $\{a\}$ for which $(d/a) = -1$ but then for a fixed b the condition on c is $(C/c) = -\epsilon(B/b)$. Consider $\sum\langle a/r\rangle$ where the summation is over all a for which $(d/a) = 1$. It follows that this is $\sum_b\sum_c\langle(cB + bC)/r\rangle$ where here the summation is over all b, but c is subject to $(C/c) = \epsilon(B/b)$. By our remarks this equals $\phi(|C|)/2\cdot\sum_b\langle bC/r\rangle$ since $r = |B|$. Now clearly $\sum\langle a/r\rangle$ where the summation is over all a for which $(d/a) = -1$ gives the same result. This proves the lemma.

Now let H be the subgroup of $LF(2,n)$ consisting of elements of the form $\pm\begin{pmatrix} a & b \\ 0 & d\end{pmatrix}$, and let L be the corresponding subfield of K. We note that L is of degree $\psi(n)$ over $\boldsymbol{C}(j)$. With the aid of Hilbert's Galois theory [6], ramification in L and K_0 will now be discussed. We note that $K \supset K_0 \supset L$.

LEMMA 3. *There are* 2^t *places of* L *over* $j = \infty$, *one place with ramification index* r *for every divisor* r *of* n. *These places of* L *are unramified in* K_0. *A place of* K_0 *which is over* $j = 0$ *or* 12^3 *is unramified in* K.

Proof. Decompose $LF(2,n)$ into left cosets by H. Now (T) is an inertia group of K over $C(j)$ at $j = \infty$ and (T) operates on this homogeneous space. Hence the space splits into orbits and if σH is a point of the space, then the multiplicity of the orbit to which σH belongs is the order of the group $(T) \cap \sigma H\sigma^{-1}$. According to Hilbert's Galois theory, there is a one-to-one correspondence between places of L over $j = \infty$ and these orbits. Furthermore, a place of L will be ramified in K with index e where e is the multiplicity of the corersponding orbit and an inertia group of K over L at this place is $H \cap \sigma^{-1}(T)\sigma$. Consider $(T) \cap \sigma H\sigma^{-1}$. This has order r if and only if it equals (T^m) where $mr = n$. But then

$$\sigma^{-1}(T^m)\sigma = \pm\begin{pmatrix} 1+cdm & d^2m \\ -c^2m & 1-cdm \end{pmatrix}$$

and this is in H if and only if $c \equiv 0 \bmod r$. However $(c, m) = 1$ since otherwise $(T) \cap \sigma H \sigma^{-1}$ would strictly contain (T^m). Therefore, using Lemma I we see that such σ's give n/r cosets σH. Since each σH belongs to an orbit consisting of n/r points, the number of orbits determined by such σ is 1. This proves the first part of the Lemma. We note also that an inertia group of K over L at this place is (T^m). Now, if P is a place of L over $j = \infty$ whose index is r, we decompose H by (T) and let (T^m) operate on this homogeneous space where $mr = n$. But clearly $(T^m) \cap \sigma(T)\sigma^{-1} = (T^m)$ for all σ in H and the second part of the lemma follows at once. Finally we decompose $LF(2,n)$ by (T) and let A operate on this homogeneous space where A is an inertia group of K over $C(j)$ at either $j = 0$ or $j = 12^3$. An inertia group of K over a place of K_0 must have the form $(T) \cap \sigma^{-1}A\sigma$. But A is conjugate to either the cyclic group generated by $\pm\begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}$ or by $\pm\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and so the only element in A which has trace ± 2 is the identity. This completes the proof of the lemma. We note that the first part of this result is contained in [7].

Now let $\{\pm \alpha_i\}$ for $i = 1, 2, \cdots, \phi(n)/2$ be a complete system of residues mod n which are relatively prime to n, and let $S_i = \pm\begin{pmatrix} \alpha_i & 0 \\ 0 & \alpha_i^{-1} \end{pmatrix}$. Let P_r be the place of L with ramification index n/r over $j = \infty$. There are $\phi(n)/2$ place $Q_{r,1}, Q_{r,2}, \cdots$ of K_0 above P_r and we denote by $P_{r,i}$ a place of K above $Q_{r,i}$ for $i = 1, \cdots, \phi(n)/2$. We can assume that $S_i(P_{r,1}) = P_{r,i}$ and consequently, if t_1 is a local parameter at $P_{r,1}$, then $S_i(t_1) = t_i$ is a local parameter at $P_{r,i}$. There is an element θ in K such that $K = K_0(\theta)$ and $\theta^n = f$ is in K_0, and if $\psi = \psi_a$ with $(a, n) = 1$ is a simple character of (T), we can also assume $T: \theta \rightarrow \psi(T)\theta$. Denote the order of θ at $P_{r,i}$ by $m_{r,i}$, and as before let $m = n/r$. Now $T^m: t_1 \rightarrow e(s_r m/n) t_1 \bmod t_1^2$ for some integer s_r and so $T^m: \theta \rightarrow e(m_{r,1}s_r m/n)\theta$. As a consequence

$$m_{r,1}s_r m \equiv am \bmod n \text{ or } m_{r,1}s_r \equiv a \bmod r.$$

We can certainly take $(s_r, n) = 1$ and then if $s_r c_r \equiv 1 \bmod n$, we finally get $m_{r,1} \equiv c_r a \bmod r$. Taking into account that $S_i(t_1) = t_i$ and $T^m S_i = S_i T \alpha_i^{-2} m$ we find that $m_{r,i} \equiv \alpha_i^2 c_r a \bmod r$ for $i = 1, 2, \cdots, \phi(n)/2$.

We now fix a differential dx in K_0, and call Ω_ψ the complex vector-space of differentials of the first kind ω in K such that $T: \omega \rightarrow \psi(T)\omega$. It is clear that $\dim \Omega_\psi$ is the multiplicity of ψ in X_0, and that $\omega \in \Omega_\psi$ if and only if

$\theta^{-1}\omega = \phi dx$ for some $\phi \in K_0$. Consequently $\dim \Omega_\psi$ is the dimension (over $\boldsymbol{C}$) of the vector-space of functions ϕ in K_0 such that $\theta\phi dx$ is a differential of the first kind in K. We denote this dimension by N_a. If we let $v_{r,i}$ be the valuation of K_0 at $Q_{r,i}$ and v be the valuation at a typical place $\mathfrak{p}$ of K_0 which is unramified in K, then the conditions on ϕ are easily seen to be

$$rv_{r,i}(\phi) + m_{r,i} + r - 1 + rv_{r,i}(dx) \geqq 0,$$
$$r(\phi) + 1/n \cdot v(f) + v(dx) \geqq 0.$$

Letting $w_{r,i}$ denote the least positive residue of $m_{r,i} \bmod r$, dividing the first condition by r and taking the integral part yield the equivalent conditions

$$v_{r,i}(\phi) + (m_{r,i} - w_{r,i})/r + 1 + v_{r,i}(dx) \geqq 0,$$
$$v(\phi) + 1/n \cdot v(f) + v(dx) \geqq 0.$$

Therefore, applying the Riemann-Roch theorem for the function field K_0 of one variable [cf. 1], we get

$$N_a = g_0 - 1 + S$$

where g_0 is the genus of K_0 and in the sum S each divisor r of n with $r > 1$ makes a contribution $\sum_{i=1}^{\phi(n)/2} \langle -\alpha_i^2 ac_r/r \rangle$. Consequently

$$N_a = g_0 - 1 + 2^{t-1} \sum_r \sum_{a_i} \langle -a_i ac_r/r \rangle$$

where the summation is over all divisors r of n and $\phi(n)/2^t$ representatives a_i of classes of type (1). Now $N_a = N_b$ if a and b are of the same type (e). Calling this multiplicity $N_{(e)}$, recalling that $(d/a) = (-1)^{s(e)}$ if $a \sim (e)$ and using Lemma 2, we get

$$\sum_{(e)} (-1)^{s(e)} N_{(e)} = -2^{t-1}(d/c_n) \sum_a (d/a) \cdot a/n$$
$$= \pm 2^{t-1} h(-n) \quad (n > 3),$$

where $h(-n)$ is the class-number of $Q(\sqrt{-n})$. The result of this section can be stated as a

THEOREM. *Suppose that $n = p_1 p_2 \cdots p_t$ where the p_i are distinct odd primes for* $\mathrm{i} = 1, 2, \cdots, t$, *that $n > 3$ and $n \equiv 3 \bmod 4$. Let X_0 be the representation of (T) on the vector space of differentials of the first kind in K, the field of modular functions of level n. Finally let N_a denote the multiplicity of the simple character ψ_a of (T) in X_0, where $(a, n) = 1$. Then $N_a = N_b$ if a and b are of the same type (e) and calling this multiplicity $N_{(e)}$, we have* $\sum_{(e)} (-1)^{s(e)} N_{(e)} = \pm 2^{t-1} h(-n)$.

3. Generalized Hecke Formula (definitive form). Let χ be a typical simple character of $LF(2,n)$ with multiplicity m_χ in X i.e. $X = \sum_\chi m_\chi \chi$. Suppose now that $\mathrm{res}_{(T)}\chi = \sum_a r_a{}^\chi \psi_a + \cdots$ where in the summation only those characters ψ_a of (T) are included for which $(a,n) = 1$. Then, with the notation of the previous section $N_a = \sum_\chi m_\chi r_a{}^\chi$. Thus every χ makes a contribution to N_a and as a consequence to $\sum_{(e)} (-1)^{s(e)} N_{(e)}$. In this section these contributions are examined. In this connection the following remark will be useful. Suppose we fix a proper subset of the set of all (e) by demanding that e_i be fixed for $1 \leqq i \leqq u$ with $u < t$ while e_j is arbitrary for $u < j \leqq t$. Call such a subset E_u. Suppose further that χ is a simple character of $LF(2,n)$ with the property that $r_a{}^\chi = 1$ if $a \sim (e)$ where $(e) \in E_u$, and $r_a{}^\chi = 0$ otherwise. Then the contribution of χ to $\sum_{(e)} (-1)^{s(e)} N_{(e)}$ is clearly $m_\chi \sum_{(e) \in E_u} (-1)^{s(e)}$. This is zero since $u < t$, and so such a χ may be omitted from our calculations. We shall also require the following simple lemma.

LEMMA. *Suppose G, G_1, G_2 are finite groups with $G = G_1 \times G_2$. Let $\phi = \phi_1\phi_2$ be a simple character of G in which ϕ_1 is principal, and let $H = H_1 \times H_2$ be a subgroup of G and $\psi = \psi_1\psi_2$ be a simple character of H in which ψ_1 is not principal. Then the multiplicity of ψ in* $\mathrm{res}_H \phi$ *is zero.*

Proof. We have

$$\begin{aligned}\sum_{h \in H} \phi(h)\psi(h^{-1}) &= \sum_{h_1 \in H_1} \sum_{h_2 \in H_2} \phi_1(h_1)\phi_2(h_2)\psi_1(h_1^{-1})\psi_2(h_2^{-1}) \\ &= \sum_{h_1 \in H_1} \phi_1(h_1)\psi_1(h_1^{-1}) \sum_{h_2 \in H_2} \phi_2(h_2)\psi_2(h_2^{-1})\end{aligned}$$

and this is zero since the first factor is zero.

It follows easily from this that if χ_i is a simple character of $SL(2,p_i)$ where $i = 1, 2, \cdots, t$ such that $\chi = \chi_1\chi_2 \cdots \chi_t$ is a character of $LF(2,n)$ and χ_i is principal for some i then the multiplicity of ψ_a in $\mathrm{res}_{(T)}\chi$ is zero if $(a,n) = 1$.

Now since we are only interested in the values of the characters on (T), it is apparent from the table at the end of the paper that the simple characters of $SL(2,p)$ are essentially of three types:

1) $\chi^{(0)}$, the principal character.

2) ϕ, $\phi(E) = p + \gamma$, $\phi(F) = (-1)^z(p+\gamma)$
$\phi(T^r) = \gamma$, $1 \leqq r \leqq p-1$,
and $\gamma = 1$, $z = \alpha$ if ϕ is $\chi^{(\alpha)}$
$\gamma = -1$, $z = \beta$ if ϕ is $\chi^{(\beta)}$
$\gamma = 0$, $z = 0$ if ϕ is $\chi^{(1)}$

3) ψ, $\psi(E) = \frac{1}{2}(p+\delta)$, $\psi(F) = \frac{1}{2}\delta(-1/p)(p+\delta)$
$\psi(T^r) = \frac{1}{2}\delta(1+\delta(r/p)g)$, $1 \leqq r \leqq p-1$,

where $\delta = \pm 1$ and $g = \pm G(p)$ where $G(p)$ is the Gaussian sum $\sum\limits_a (a/p)e(a/p)$. The ψ's are the characters of class (G) of $SL(2,p)$.

Now if χ is a simple character of $LF(2,n)$, then as has been remarked $\chi = \chi_1\chi_2 \cdots \chi_t$ where χ_i is a simple character of $SL(2,p_i)$ for $1 \leqq i \leqq t$ and $\prod\limits_{i=1}^{t} \chi_i(E_i) = \prod\limits_{i=1}^{t} \chi_i(F_i)$. We wish to compute $\sum\limits_r \chi(T^r)\psi_a(T^{-r})$ when $(a,n) = 1$. By the lemma we can restrict ourselves to the case when each χ_i is a ϕ_i or a ψ_i where ϕ and ψ have the meanings assigned above and the subscript 'i' refers to $SL(2,p_i)$. By a re-arrangement if necessary we can assume $\chi = \psi_1 \cdots \psi_m\phi_{m+1} \cdots \phi_t$. To make the computation we sum over the complete residue system mod n defined by $r = r_1s_1 + r_2s_2 + \cdots + r_ts_t$ where $s_i = n/p_i$ and $0 \leqq r_i \leqq p_i - 1$ for $i = 1, 2, \cdots, t$. Summing first over those r for which $(r,n) = 1$ gives

$$1/2^m \cdot \prod_{i=1}^{m} \delta_i \cdot \prod_{j=m+1}^{t} \gamma_j \cdot \sum_{(r,n)=1} \Big(\prod_{k=1}^{m} [1 + \delta_k(r/p_k)g_k]\Big) e(-ar/n).$$

Multiplying out the last product in this expression, summing over r and using the definition of Gaussian sum we get

$$1/2^m \cdot (-1)^t \cdot \prod_{i=1}^{m} \delta_i \cdot \prod_{j=m+1}^{t} \gamma_j \cdot \prod_{k=1}^{m} [1 - \delta_k(-as_k/p_k)g_kG(p_k)].$$

We put $b_i = (-as_i/p_i)g_iG(p_i)$. Our expression then becomes

$$1/2^m \cdot (-1)^t \cdot \prod_{j=m+1}^{t} \gamma_j \prod_{k=1}^{m} (\delta_k - b_k). \tag{1}$$

Summing then over those r for which $(r,n) > 1$,

$$(r,n) = p_{u+1} \cdots p_m \cdot p_{v+1} \cdots p_t$$

say, gives by a calculation similar to the above

$$1/2^m \cdot (-1)^u \prod_{i=1}^{u} (\delta_i - b_i) \prod_{j=u+1}^{m} (p_j + \delta_j) \cdot (-1)^{v-m} \prod_{k=u+1}^{v} \gamma_k \prod_{l=v+1}^{t} (p_l + \gamma_l). \tag{2}$$

Finally, summing (1) and all expressions of the form of (2) gives

$$1/2^m \cdot \prod_{i=1}^{m} [(p_i + \delta_i) - (\delta_i - b_i)] \cdot \prod_{j=m+1}^{t} [(p_j + \gamma_j) - \gamma_j].$$

Consequently

$$\sum_r \chi(T^r)\psi_a(T^{-r}) = 1/2^m \cdot \prod_{i=1}^{m} (p_i + b_i) \cdot \prod_{j=m+1}^{t} p_j.$$

Now $g_i = e_iG(p_i)$ with $e_i = \pm 1$, and $G(p_i^2 = (-1/p_i)p_i$ for $1 \leqq i \leqq t$

and so our sum is $n/2^m \cdot \prod_{i=1}^{m} [1 + e_i(s_i/p_i)(a/p_i)]$. It is clear that this is zero unless $(a/p_i) = e_i(s_i/p_i)$ for $1 \leqq i \leqq m$, and then our sum equals n. And so, by a remark made in the beginning of this section χ will contribute nothing to $\sum_{(e)} (-1)^{s(e)} N_{(e)}$ if $m < t$. However when $m = t$, χ will contain ψ_a once if $(a/p_i) = e_i(s_i/p_i)$ for every i, and not at all otherwise.

The preceding argument shows that the only characters of $LF(2n)$ which contribute to $\sum_{(e)} (-1)^{s(e)} N_{(e)}$ are those of the form $\chi = \psi_1 \psi_2 \cdots \psi_t$, i.e. the characters of class (G). Now $\psi_i(T^r) = \frac{1}{2} \cdot \delta_i[1 + \delta_i(r/p_i) g_i]$ where $\delta_i = \pm 1$ and $g_i = \pm G(p_i)$ for $1 \leqq i \leqq t$. We shall say that χ is of type $(e) = (e_1, e_2, \cdots, e_t)$ when $g_i = e_i G(p_i)$ for $1 \leqq i \leqq t$. We have shown that if χ is of type (e), then the multiplicity of ψ_a in $\mathrm{res}_{(T)}\chi$ is one if a is of type $(e_1', e_2', \cdots, e_t')$ where $e_i' = (s_i/p_i) e_i$ for $1 \leqq i \leqq t$ and is zero otherwise. It is easily seen that $LF(2, n)$ has 2^{t-1} characters which are of type (e) for each (e). We denote these characters by $\chi_{(e)}{}^i$, and their multiplicities in X by $n_{(e)}{}^i$ where $1 \leqq i \leqq 2^{t-1}$. Then

$$\chi_0 = \sum_{i,(e)} n_{(e)}{}^i \, \mathrm{res}_{(T)} \chi_{(e)}{}^i = \sum_{i,(e)} n_{(e)}{}^i \sum_{a \sim (e')} \psi_a.$$

Consequently $N_{(e')} = \sum_{i=1}^{2^{t-1}} n_{(e)}{}^i$ and

$$\sum_{(e')} (-1)^{s(e')} N_{(e')} = \prod_{i=1}^{t} (s_i/p_i) \sum_{(e)} \sum_{i=1}^{2^{t-1}} (-1)^{s(e)} n_{(e)}{}^i$$

and so by the theorem in the previous section,

$$\chi_0 = \sum_{(e)} \sum_{i=1}^{2^{t-1}} (-1)^{s(e)} n_{(e)}{}^i = \pm 2^{t-1} h(-n).$$

We can therefore state our

MAIN THEOREM. *Let* $n = p_1 p_2 \cdots p_t$ *where for* $1 \leqq i \leqq t$ *the* p_i *are distinct odd primes,* $n > 3$ *and* $n \equiv 3 \bmod 4$. *Let* X *be the representation of* $LF(2, n)$ *on the vector-space of differentials of the first kind in* K, *the field of modular functions of level* n. *Finally let* $\chi_{(e)}{}^i$, *where* $1 \leqq i \leqq 2^{t-1}$, *be the characters of* $LF(2, n)$ *which are of class* (G) *and type* (e), *and* $n_{(e)}{}^i$ *the multiplicity of* $\chi_{(e)}{}^i$ *in* X.

Then

$$\sum_{i,(e)} (-1)^{s(e)} n_{(e)}{}^i = \pm 2^{t-1} h(-n),$$

where $s(e)$ *denotes the number of members of* $(e_1, e_2, \cdots, e_t)$ *which are negative.*

Modified Table of Characters of $SL(2,p)$ where p is an odd prime.

	$\chi^{(0)}$	$\chi^{(1)}$	$\chi^{(\alpha)}$	$\chi^{(\beta)}$	$\psi_\pm$	$\psi_\pm'$
E	1	p	$p+1$	$p-1$	$\frac{1}{2}(p+1)$	$\frac{1}{2}(p-1)$
F	1	p	$(-1)^\alpha(p+1)$	$(-1)^\beta(p-1)$	$\epsilon/2(p+1)$	$-\epsilon/2(p-1)$
P	1	0	1	-1	$\frac{1}{2}(1\pm\sqrt{\epsilon p})$	$\frac{1}{2}(-1\pm\sqrt{\epsilon p})$
Q	1	0	1	-1	$\frac{1}{2}(1\mp\sqrt{\epsilon p})$	$\frac{1}{2}(-1\mp\sqrt{\epsilon p})$

Explanation. $E=\begin{pmatrix}1&0\\0&1\end{pmatrix}$, $F=-\begin{pmatrix}1&0\\0&1\end{pmatrix}$, $P=\begin{pmatrix}1&0\\1&1\end{pmatrix}$, $Q=\begin{pmatrix}1&0\\u&1\end{pmatrix}$, where u is a quadratic non-residue $\operatorname{mod} p$; $\epsilon=(-1/p)$; $1\leqq\alpha\leqq\frac{1}{2}(p-3)$ and $1\leqq\beta\leqq\frac{1}{2}(p-1)$. We note that P is conjugate to $\begin{pmatrix}1&-1\\0&1\end{pmatrix}$ and Q is conjugate to $\begin{pmatrix}1&-u\\0&1\end{pmatrix}$. Furthermore T^r and T^s are conjugate if and only if $(r/p)=(s/p)$.

UNIVERSITY COLLEGE, GALWAY.

REFERENCES.

[1] C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable,* American Mathematical Society (1951).

[2] ——— and A. Weil, "Über das Verhalten der Integral 1. Gattung bei Automorphismen des Funktionenkörpers," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität,* pp. 358-361.

[3] E. Hecke, "Über ein Fundamentalproblem aus der Theorie der Elliptischen Modulfunktionen," *Collected Works* (1959), pp. 525-547.

[4] ———, "Über das Verhalten der Integrale 1. Gattung bei Abbildungen," *ibid.,* pp. 548-558.

[5] ———, *Vorlesungen über die Theorie der Algebraischen Zahlen,* Chelsea (1948).

[6] D. Hilbert, "Die Theorie der Algebraischen Zahlkörper," *Jahresbericht der Deutschen Mathematiker Vereinigung,* vol. 4 (1897), pp. 177-546.

[7] J. Igusa, "Fibre systems of Jacobian varieties" (III. Fibre Systems of Elliptic Curves), *American Journal of Mathematics,* vol. 81 (1959), pp. 453-476.

[8] R. Kawai, "A note on Riemann-Roch-Weil's theorem," *Memoirs of the College of Science, University of Kyoto,* Series A, vol. 27 (1952), pp. 123-131.

[9] F. Klein, "Zur Theorie der Elliptischen Modulfunktionen," *Collected Works,* vol. 3 (1923), pp. 169-178.

[10] S. Lang, *Introduction to Algebraic Geometry,* Interscience (1958).

[11] F. Murnaghan, *Theory of Group Representations,* The Johns Hopkins University (1938).

[12] H. Rohrbach, "Die Charaktere der binären Kongruenzgruppen mod p^2," *Schriften des Mathematischen Seminars und des Instituts für angewandte Mathematik der Universität Berlin,* vol. 1 (1932-33), pp. 33-94.

SOLVABLE FUNCTIONS FOR THE DIRICHLET PROBLEM.*

By Ernest C. Schlesinger.

1. Introduction. On a Riemann surface we shall consider a region Ω, regular for the Dirichlet problem, whose boundary Γ is compact. Moreover, we assume that the maximum principle is valid in Ω. Under these conditions the Perron method ([5]; [1], Section 11) solves the Dirichlet problem. We recall the essence of this method. If f is a continuous function on Γ we form the family $\mathfrak{V}(f)$ of subharmonic functions in Ω whose behavior near Γ is dominated by f in the sense that for each $v \in \mathfrak{V}(f)$ and each point $\zeta \in \Gamma$,

$$\limsup_{z \to \zeta} v(z) \leqq f(\zeta). \tag{1}$$

The function u_f defined by

$$u_f(z) = \operatorname{lub}\{v(z) \mid v \in \mathfrak{V}(f)\} \tag{2}$$

is then the solution of the Dirichlet problem with boundary values f:

$$\hat{u}_f(z) = \begin{cases} u_f(z), z \in \Omega \\ f(z), z \in \Gamma \end{cases}$$

is continuous in $\Omega \cup \Gamma$ and harmonic in Ω ([1], Section 11).

The same procedure can be applied to functions that are not necessarily continuous on Γ. We shall restrict our considerations to functions f that are bounded, however. If $\mathfrak{V}(f)$ is defined as above, the function u_f will again be harmonic, and, by virtue of our assumption of regularity, it will satisfy the inequality

$$\liminf_{\zeta \to \zeta^*} f(\zeta) \leqq \liminf_{z \to \zeta^*} u_f(z) \leqq \limsup_{z \to \zeta^*} u_f(z) \leqq \limsup_{\zeta \to \zeta^*} f(\zeta), \tag{3}$$

$(z \in \Omega, \zeta, \zeta^* \in \Gamma)$, at each boundary point ζ^* ([1], p. 140).

It is clear (see below, Section 2) that u_f is superadditive in its dependence on f:

$$u_f + u_g \leqq u_{f+g}. \tag{4}$$

The question arises under what conditions the opposite inequality holds. The

---

* Received February 8, 1962.

author recently proved directly that if f and g are both lower semi-continuous on Γ then

$$u_f + u_g = u_{f+g}. \tag{5}$$

A sketch of the proof is given in Section 2; the lower semi-continuity of f and g enters significantly in the reasoning. This result led to a search for a "reasonable" counterexample outside the class of lower semi-continuous functions. The purpose of the remainder of the paper is to show that (5) holds for all functions that are measurable in a sense to be made precise in the course of the argument. A counterexample of the type just envisaged must, therefore, involve non-measurable functions.

2. Additivity in the lower semi-continuous case. If f and g are bounded functions on Γ, and if $v_1 \in \mathfrak{V}(f)$ and $v_2 \in \mathfrak{V}(g)$, then $v_1 + v_2$ is subharmonic in Ω and

$$\begin{aligned}\limsup_{z\to\zeta}(v_1+v_2)(z) &\leqq \limsup_{z\to\zeta} v_1(z) + \limsup_{z\to\zeta} v_2(z)\\ &\leqq f(\zeta)+g(\zeta).\end{aligned}$$

Hence, $v_1 + v_2 \in \mathfrak{V}(f+g)$, and, therefore, $v_1 + v_2 \leqq u_{f+g}$. From this inequality, (4) follows in a standard way.

We now assume that f and g are not merely bounded but also lower semi-continuous on Γ. Let $v \in \mathfrak{V}(f+g)$ and $\zeta^* \in \Gamma$, and consider

$$\limsup_{z\to\zeta^*}(v - u_f - u_g)(z), \quad (z \in \Omega).$$

This is at most

$$\limsup_{z\to\zeta^*} v(z) - \liminf_{z\to\zeta^*} u_f(z) - \liminf_{z\to\zeta^*} u_g(z),$$

which, in view of (3), is less than or equal to

$$\begin{aligned}(f+g)(\zeta^*) &- \liminf_{\zeta\to\zeta^*} f(\zeta) - \liminf_{\zeta\to\zeta^*} g(\zeta)\\ &= f(\zeta^*) - \liminf_{\zeta\to\zeta^*} f(\zeta) + g(\zeta^*) - \liminf_{\zeta\to\zeta^*} g(\zeta),\end{aligned}$$

$(\zeta \in \Gamma)$. By virtue of the lower semi-continuity of f and g, the last expression is at most 0. From the maximum principle we conclude that for all z in Ω,

$$v(z) \leqq u_f(z) + u_g(z).$$

Hence,

$$u_{f+g} = \operatorname{lub}\{v(z) \mid v \in \mathfrak{V}(f+g)\} \leqq u_f(z) + u_g(z).$$

This relation, together with (4), yields the desired result (5).

3. The classes $\mathcal{S}$ and $\mathcal{L}$. We continue to restrict our considerations to bounded functions on Γ. Let f be such a function. We dualize the Perron procedure ([5], [8]). Namely, we form $\mathcal{W}(f)$, the family of all superharmonic functions on Ω, which, for each boundary point $\zeta \in \Gamma$, satisfy the condition

$$\liminf_{z \to \zeta} w(z) \geqq f(\zeta) \qquad (z \in \Omega).$$

The function U_f, defined in Ω by

$$U_f(z) = \operatorname{glb}\{w(z) \mid w \in \mathcal{W}(f)\},$$

is, once more, harmonic. Also, U_f in place of u_f satisfies the inequality (3). Moreover,

$$u_f(z) \leqq U_f(z) \qquad (z \in \Omega), \tag{6}$$

since it follows from the maximum principle that $v \leqq w$ for all v and w belonging to $\mathcal{V}(f)$ and $\mathcal{W}(f)$, respectively. Equality can hold in (6) at a point $z_0 \in \Omega$ if and only if it holds throughout the region Ω.

We now introduce two classes of functions on Γ. The first of these is

$$\mathcal{S} = \{f \mid u_f = U_f\}.$$

To define the other, we must review briefly Daniell's procedure ([2]) of extending an elementary integral. We shall follow the exposition of this procedure given by Loomis ([4], Section 12).

Let us fix a point $z_0 \in \Omega$ and set

$$I_0(f) = I_0(f; z_0) = u_f(z_0).$$

The functional I_0 and the family $\mathcal{C} = \mathcal{C}(\Gamma)$ of continuous functions on Γ satisfy the following conditions: If f, g, and f_n denote functions of $\mathcal{C}$ and if α is a real number, then

a. $I_0(f+g) = I_0(f) + I_0(g)$,

b. $I_0(\alpha f) = \alpha I_0(f)$,

c. If $f \geqq 0$ *then* $I_0(f) \geqq 0$, *and*

d. If the sequence $\{f_n\}$ *decreases (pointwise) monotonically to the constant function* 0, *then the sequence of real numbers* $\{I_0(f_n)\}$ *decreases monotonically to* 0.

The first three properties are immediate consequences of the definitions of I_0 and of (3); the fourth follows from Dini's theorem (e.g. [4], Section 16A, Lemma 1) which applies since Γ is compact. These conditions allow us

([4], Section 12) to extend $\mathscr{B}$ and I_0 to $\mathscr{L} = \mathscr{L}(z_0)$ and $I = I(\cdot\,; z_0)$, respectively, where $\mathscr{L}$ is the Lebesgue class generated by $\mathscr{B}$ and I_0, and I is the corresponding integral.

Specifically, let $\mathfrak{U}$ denote the family of lower semi-continuous "summable" functions: $\psi \in \mathfrak{U}$ if and only if there is a monotonically increasing sequence $\{f_n\}$ of functions of $\mathscr{B}$ that converges to ψ and for which $I_0(\psi)$, defined as $\lim I_0(f_n)$, is finite. (Because of property *c.*, above, this limit exists in any event, but it may be $+\infty$.) Then we say that the function f on Γ belongs to $\mathscr{L}$ if and only if for every $\epsilon > 0$ there are functions ϕ and ψ, with $-\phi$ and ψ in $\mathfrak{U}$, $\phi \leqq f \leqq \psi$, and $I_0(\psi) - I_0(\phi) < \epsilon$. If f is "summable," i.e. if f is in $\mathscr{L}$, we define the Lebesgue integral I of f by

$$\begin{aligned} I(f) = I(f; z_0) &= \operatorname{glb}\{I_0(\psi) \,|\, f \leqq \psi \textit{ and } \psi \in \mathfrak{U}\} \\ &= \operatorname{lub}\{I_0(\phi) \,|\, \phi \leqq f \textit{ and } -\phi \in \mathfrak{U}\}. \end{aligned}$$

$I(f)$ is then a (finite) real number.

Clearly, both I and $\mathscr{L}$ depend on z_0 because of the manner in which they are defined. However, our principal result is the following:

THEOREM. *For all* $z_0 \in \Omega$, $\mathscr{L}(z_0) = \mathscr{S}$.

Consequently, the dependence of $\mathscr{L}$ on z_0 is spurious. The proof is contained in the next two sections.

4. Two lemmas.

LEMMA 1. *For all* $z \in \Omega$, $\mathfrak{U} \subset \mathscr{S} \cap \mathscr{L}(z)$.

Proof. This proof is an adaptation of one used by Pfluger ([6], p. 99) for essentially the same purpose. The method seems to go back to Daniell ([2], p. 283). Since $\mathfrak{U} \subset \mathscr{L}(z)$ by the Daniell construction, it suffices to show that $\psi \in \mathfrak{U}$ implies $\psi \in \mathscr{S}$. Let $\{f_n\}$, $(n = 1, 2, \cdots)$, be a monotonically increasing sequence of functions of $\mathscr{B}$ that converges to $\psi \in \mathfrak{U}$. If z is a point of Ω and n is a positive integer,

$$I_0(f_n; z) = u_{f_n}(z) = U_{f_n}(z),$$

and $\lim\limits_{z \to \zeta} I_0(f_n; z) = f_n(\zeta) \leqq \psi(\zeta)$, by inequality (3). Hence, for $n \geqq 1$, $I_0(f_n; \cdot) \in \mathfrak{V}(\psi)$, so that

$$I_0(f_n; z) \leqq \operatorname{lub}\{v(z) \,|\, v \in \mathfrak{V}(\psi)\} = u_\psi(z),$$

and

$$I_0(\psi; z) = \lim I_0(f_n; z) \leqq u_\psi(z). \tag{7}$$

To prove $I_0(\psi) \geqq U_\psi$, we fix $z_0 \in \Omega$ and set $g_n = f_{n+1} - f_n$, $(n = 0, 1, 2, \cdots)$, where we agree that $f_0 \equiv 0$. The functions g_n belong to $\mathcal{B}$, and $g_n \geqq 0$, at least for $n \geqq 1$. Hence, for each $\epsilon > 0$ there is a function $w_n \in \mathcal{W}(g_n)$ such that

$$w_n(z_0) < I_0(g_n; z_0) + \epsilon/2^{n+1}, \qquad (n = 0, 1, 2, \cdots), \tag{8}$$

where, at least for $n \geqq 1$, $w_n \geqq 0$. We sum the inequalities (8) from $n = 0$ to $n = p - 1$:

$$\begin{aligned}\sum_{n=0}^{p-1} w_n(z_0) &< \sum_{n=0}^{p-1} I_0(g_n; z_0) + \epsilon(1 - 2^{-p}) \\ &\leqq I_0(f_p; z_0) + \epsilon \leqq I_0(\psi; z_0) + \epsilon < +\infty,\end{aligned}$$

and we see that the series $w(z) = \sum_{n=0}^{\infty} w_n(z)$ converges absolutely at z_0. It is, thus, the limit of an increasing sequence of superharmonic functions that is bounded above at one point. Consequently, w is itself superharmonic in Ω ([7], p. 14). For each $p \geqq 1$ and all $\zeta \in \Gamma$ we also have

$$\begin{aligned}\liminf_{z\to\zeta} w(z) \geqq \liminf_{z\to\zeta} \sum_{n=0}^{p-1} w_n(z) &\geqq \sum_{n=0}^{p-1} \liminf_{z\to\zeta} w_n(z) \\ &\geqq \sum_{n=0}^{p-1} g_n(\zeta) = f_p(\zeta).\end{aligned}$$

Hence, $\liminf_{z\to\zeta} w(z) \geqq \lim f_p(\zeta) = \psi(\zeta)$, so that $w \in \mathcal{W}(\psi)$. We conclude that

$U_\psi(z_0) \leqq w(z_0) \leqq I_0(\psi; z_0) + \epsilon$, and, since ϵ is arbitrary,

$$U_\psi(z_0) \leqq I_0(\psi; z_0). \tag{9}$$

Inequalities (7) and (9), together with the remark about equality in (6), show that $\psi \in \mathcal{S}$.

COROLLARY. —$\mathcal{U} = \{\phi \mid -\phi \in \mathcal{U}\} \subset \mathcal{S} \cap \mathcal{L}(z)$ *for all* $z \in \Omega$.

The proof is a dual of the preceding one.

LEMMA 2. *Let* $v \in \mathcal{V}(f)$, *and define* $\bar{v}$ *as a function on* Γ *by* $\bar{v}(\zeta) = \limsup_{z\to\zeta} v(z)$. *Then* $v \in \mathcal{V}(\bar{v})$, $\bar{v} \in \mathcal{U}$, *and* $I_0(\bar{v}; \cdot) \in \mathcal{V}(f)$.

Proof. The first inclusion is obvious from the definition of $\bar{v}$ and the second one holds because $\bar{v}$ is a limit superior. ([3], p. 308, has a similar result whose proof can be adapted immediately to the present situation.) Using Lemma 1 we conclude that $I_0(\bar{v}; \cdot)$ is harmonic and, *a fortiori*, subharmonic in Ω. Next, the inequality (3), the definition of $\bar{v}$, and the condition that $v \in \mathcal{V}(f)$ yield that for each $\zeta \in \Gamma$:

$$\limsup_{z \to \zeta} I_0(\bar{v}; z) \leqq \bar{v}(\zeta) = \limsup_{z \to \zeta} v(z) \leqq f(\zeta),$$

and the third inclusion is established.

COROLLARY. *Let* $w \in \mathcal{W}(f)$ *and set* $\underline{w}(\zeta) = \liminf_{z \to \zeta} w(z)$ *for each point* $\zeta \in \Gamma$. *Then* $w \in \mathcal{W}(\underline{w})$, $-\underline{w} \in \mathcal{U}$, *and* $I_0(\underline{w}; \cdot) \in \mathcal{W}(f)$.

5. The main result.

THEOREM. *For each* $z_0 \in \Omega$, $\mathcal{L}(z_0) = \mathcal{S}$.

Proof. 1. If $f \in \mathcal{L}(z_0)$, then for every positive number ϵ there are functions ϕ and ψ, such that $-\phi$ and ψ belong to $\mathcal{U}$, $\phi \leqq f \leqq \psi$, and

$$0 \leqq I_0(\psi; z_0) - I_0(\phi; z_0) < \epsilon.$$

The functions ϕ and ψ depend on f, z_0, and ϵ. By Lemma 1, $I_0(\phi; \cdot)$ and $I_0(\psi; \cdot)$ are harmonic functions. Inequality (3) and the membership of $-\phi$ and ψ in $\mathcal{U}$ yield, moreover, that

$$\limsup_{z \to \zeta^*} I_0(\phi; z) \leqq \limsup_{\zeta \to \zeta^*} \phi(\zeta) \leqq \phi(\zeta^*) \leqq f(\zeta^*)$$

and

$$\liminf_{z \to \zeta^*} I_0(\psi; z) \geqq \liminf_{\zeta \to \zeta^*} \psi(\zeta) \geqq \psi(\zeta^*) \geqq f(\zeta^*),$$

so that $I_0(\phi; \cdot) \in \mathcal{V}(f)$ and $I_0(\psi; \cdot) \in \mathcal{W}(f)$. Thus,

$$I_0(\phi; z_0) \leqq u_f(z_0) \leqq U_f(z_0) \leqq I_0(\psi; z_0) < I_0(\phi; z_0) + \epsilon.$$

The desired conclusion, $u_f(z_0) = U_f(z_0)$, now follows from the arbitrariness of ϵ. Therefore, $f \in \mathcal{S}$, and we have shown that for each $z_0 \in \Omega$, $\mathcal{L}(z_0) \subset \mathcal{S}$.

2. If $f \in \mathcal{S}$ then $u_f = U_f$. We fix $z_0 \in \Omega$ and $\epsilon > 0$. There exist functions $v \in \mathcal{V}(f)$ and $w \in \mathcal{W}(f)$, depending on f, z_0, and ϵ, with

$$v(z_0) > u_f(z_0) - \tfrac{1}{2}\epsilon \text{ and } w(z_0) < U_f(z_0) + \tfrac{1}{2}\epsilon.$$

We form the functions $\bar{v}$ and $\underline{w}$ of Lemma 2 and its corollary, and we find that

$$\begin{aligned} u_f(z_0) - \tfrac{1}{2}\epsilon &< v(z_0) \leqq I_0(\bar{v}; z_0) \leqq u_f(z_0) \\ &\leqq U_f(z_0) \leqq I_0(\underline{w}; z_0) \leqq w(z_0) < U_f(z_0) + \tfrac{1}{2}\epsilon = u_f(z_0) + \tfrac{1}{2}\epsilon. \end{aligned}$$

Our conclusion is that $I_0(\underline{w}; z_0) - I_0(\bar{v}; z_0) < \epsilon$. But $-\underline{w}$ and $\bar{v}$ belong to

$\mathcal{U}$, so that $f \in \mathcal{L}(z_0)$. Hence, again for each $z_0 \in \Omega$, $\mathcal{S} \subset \mathcal{L}(z_0)$. Therefore, $\mathcal{S} = \mathcal{L}(z_0)$ for each $z_0 \in \Omega$, as announced.

COROLLARY. *$\mathcal{L}(z)$ is independent of $z \in \Omega$.*

This holds, since the same is true of $\mathcal{S}$. Of course, the values of the integrals $I(\cdot\,; z)$ as functionals on $\mathcal{S} = \mathcal{L}(z)$ do depend on the point z.

Remark. A counterexample to the result of Section 2 is now found quite easily. Let f be a bounded non-measurable function on Γ. Then our theorem implies that $u_f < U_f$. Setting $g = -f$, we have

$$u_f + u_g = u_f + u_{-f} = u_f - U_f < 0 = u_0 = u_{f+g}.$$

WESLEYAN UNIVERSITY.

REFERENCES.

[1] L. V. Ahlfors and L. Sario, *Riemann Surfaces*, Princeton, 1960.

[2] P. J. Daniell, "A general form of integral," *Annals of Mathematics*, vol. 19 (1917-18), pp. 279-294.

[3] E. W. Hobson, *The theory of functions of a real variable and the theory of Fourier's series*, vol. I, 3rd ed., Cambridge, 1927.

[4] L. H. Loomis, *An introduction to abstract harmonic analysis*, New York, 1953.

[5] O. Perron, "Eine neue Behandlung der ersten Randwertaufgabe für $\Delta u = 0$," *Mathematische Zeitschrift*, vol. 18 (1923), pp. 41-54.

[6] A. Pfluger, *Theorie der Riemannschen Flächen*, Berlin, 1957.

[7] T. Radó, "Subharmonic functions," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. V, 1 (1937); reprinted, New York, 1949.

[8] N. Wiener, "Note on a paper of O. Perron," *Journal of Mathematics and Physics*, vol. 4 (1925), pp. 21-32.

SOLUTIONS OF SOME PROBLEMS OF DIVISION.*

Part V. Hyperbolic Operators.

By LEON EHRENPREIS.[1]

---

1. Introduction. The classical theory of linear hyperbolic differential operators is to a large extent concerned with Cauchy's problem. However, it is possible to give a condition which is equivalent to hyperbolicity for linear partial differential operators and which has meaning for general convolution equations and even for overdetermined systems of such equations. There is also a connection between this general hyperbolicity and Wiener-Hopf problems for several variables.

Let $R(C)$ denote real (complex) euclidean space of dimension $n+1$ with coordinates $t, x, x = (x_1, \cdots, x_n)$ $(s, z, z = (z_1, \cdots, z_n))$. Denote by E the space of indefinitely differentiable functions on R with the usual Schwartz topology. For any $b > 0$, call E_b the space of functions which are defined and C^∞ for all x and for $|t| < b$; E_b is given the Schwartz topology. Let E', E'_b denote the respective duals of E, E_b and let $\hat{E}'$, $\hat{E}'_b$ be their Fourier transforms. Let $S \in E'$ and suppose that the support of S is contained in $|t| < b_0$ where $b_0 < b$. Call $J(s, z)$ the Fourier transform of S for (s, z) the dual variables to (t, x) and call V the set of zeros of J.

Definition. S is called *hyperbolic in* t if there is a $b > b_0$ and a b_1 so that for any $f \in E_b$ with $S * f = 0$ on $|t| < b - b_0$ there is a unique $g \in E$ such that $S * g = 0$ and $g = f$ on $|t| < b - b_1$. Moreover the map $f \rightarrow g$ should be continuous from the topology of E_b into the topology of E.

We shall prove

THEOREM 1. *Suppose S is invertible* (see [4]). *A necessary and sufficient condition that S be hyperbolic in t is that there exist a $c > 0$ so that*

$$(1) \qquad |Is| \leqq c(1 + |Iz| + \log(1 + |s| + |z|))$$

for $(s, z) \in V$.

(I denotes the imaginary part.)

Theorem 1 together with a theorem of Gårding (see [8]) show that for

---

* Received March 2, 1962.

[1] This work was sponsored by the Army Research Office (DURHAM).

linear partial differential operators with constant coefficients our concept of hyperbolicity is the same as the usual one defined by Cauchy's problem being well-posed for C^∞ data.

From Theorem 1 we can deduce: Let $n = 0$ and let S be a differential-difference operator; then S is hyperbolic. (The analogous result for $n > 0$ is false.)

The essential part of the proof consists in the construction of a suitable fundamental solution for S. This is carried out in Section 2. In Section 3 we prove Theorem 1 using this fundamental solution, and we deduce the above result on differential-difference operators. An example is given to show that the hypothesis "S is invertible" cannot be eliminated.

Theorem 1 for $n = 0$ can be deduced from the expansion theorem for mean periodic functions given in [4]. For many S the result, as well as an analogous one for overdetermined systems can be deduced by other methods (see [7]) which do not seem to apply to all invertible S.

It is also possible to generalize the results by replacing the spaces E, E_b by other spaces, but we shall not discuss this here. (See [7] for a general discussion of this type of problem.)

We could define also the notion of semi-hyperbolicity: S is called *semi-hyperbolic in* $t > 0$ (resp. $t < 0$) if there exists $b > b_0$ and b_1 so that for any $f \in E_b$ which satisfies $S * f = 0$ on $|t| < b - b_0$ there is a unique $g \in E^{\pm}$ (the space of C^∞ functions in $t > 0$ (resp. $t < 0$)) with $S * g = 0$ for $t > b_0$ (resp. $t < -b_0$) such that $g = f$ on $t > b_1$ (resp. $t < -b_1$). Moreover $f \rightarrow g$ is continuous from E_b into $E^{\pm}$. In analogy to Theorem 1 we have

THEOREM 1′. *Let S be invertible. A necessary and sufficient condition that S be semi-hyperbolic for $t > 0$ (resp. $t < 0$) is that there be a $c > 0$ so that*

$$(1') \qquad \begin{gathered} Is \leqq c(1 + |Iz| + \log(1 + |s| + |z|) \\ (\text{resp. } Is \geqq c(1 + |Iz| + \log(1 + |s| + |z|)) \end{gathered}$$

for $(s, z) \in V$.

The proof of Theorem 1′ follows the same lines as that of Theorem 1 and so will be omitted.

It is also of interest to introduce a concept quasi-hyperbolicity by modifying the definition of hyperbolicity by not requiring that g be unique nor that g depend continuously on f. However in general I do not know how to characterize those S which are quasi-hyperbolic except for the case of partial differential operators. The methods are much different from those of this paper and will not be discussed here. See [7] where quasi-semi-hyperbolicity is discussed for overdetermined systems. (In [7] this concept is called "semi-hyperbolicity.")

In Section 4 we study a generalized Cauchy problem for S in case S is a partial differential operator which is hyperbolic in t. Suppose S is of order m. The usual Cauchy problem consists of determining $f \in E$ satisfying $S * f = 0$ from a knowledge of $f(0,x)$ and $\dfrac{\partial^j f}{\partial x^j}(0,x)$ for $j = 1, 2, \cdots, m-1$. Our generalization is as follows: Let $r \leqq m$, let $t_1 < t_2 < \cdots < t_r$, and let $\{\theta_{jk}\}_{\substack{j=1,\cdots,a_k \\ k=1,\cdots,r}}$ be given linear partial differential operators with constant coefficients with $\sum a_k = m$. Then we want to determine $f \in E$ satisfying $S * f = 0$ from $\{\partial_{jk} f(t_k, x)\}$. We shall show that, under a suitable independence condition on the $\{\partial_{jk}\}$, the generalized Cauchy problem has existence but usually there is no uniqueness. An even greater generalization of the Cauchy problem is considered.

In Section 5 we drop the assumption that S be a differential operator and we study an analog of the Cauchy problem for hyperbolic S. The "Cauchy data" must, of course, be given on a set $|t| \leqq b$ for suitable b. For $n = 0$ the Cauchy problem can be shown to be closely related to a Wiener-Hopf problem, so the solution of the Cauchy problem represents a natural extension of the Wiener-Hopf results to higher dimensions.

2. Construction of a fundamental solution.

Let S be invertible. By Theorem 2.2 of [4] $J(s,z)$ is slowly decreasing. Thus there is an $a > 0$ so that for any real (s,z) there is a real (s',z') with

$$|s - s'| + |z - z'| < a \log(1 + |s| + |z|) \tag{2}$$

$$|J(s',z')| \geqq a(1 + |s| + |z|)^{-a} \tag{3}$$

Suppose the zeros V of J satisfy (1). Then for z fixed, z real, the zeros s of $J(s,z)$ lie "close to" the real axis. We should like to define fundamental solutions for J by using $1/J(s,z)$ on curves of the form $\Gamma^{z\pm} = \Gamma_z^{\pm} = \Gamma_z^{\pm}(c')$:

$$Is = \pm c'(1 + \log(1 + |z| + |s|)), \tag{4}$$

where $c' > c$ so all the zeros of $J(s,z)$ lie between $\Gamma_z^{\pm}$. Then we could define two fundamental solutions $H^{\pm}$ for S by

$$H^{\pm}(t,x) = \int e^{-ix\cdot z}\, dz \int_{\Gamma_z^{\pm}} e^{-its}\, ds/J(s,z) \tag{5}$$

where dz is the element of volume in the real part of complex n space, that is, for any C^∞ function f of compact support,

$$(5') \qquad H^{\pm} \cdot f = \int e^{-ix \cdot z}\, dz \int_{\Gamma_z^{\pm}} F(s,z)\, ds/J(s,z)$$

where F is the Fourier transform of f.

We must know that the integrals defined by (5) converge in the topology of the space of distributions. For this we need

LEMMA 1. *We can find an* $a' > 0$ *depending only on* a *and the exponential type of* J *so that for every* $s \in \Gamma_z^{\pm}(4(a+c))$ *we have*

$$(6) \qquad |J(s,z)| \geqq (a')^{-1}(1 + |s| + |z|)^{-a'}.$$

Proof. For any set of points $\gamma = \{(s,z)\}$ we shall say that J *is large on* γ if (6) holds for $(s,z) \in \gamma$ with a constant a' depending only on a, n, and the exponential type of J.

Let z^0 be real and $s^0 \in \Gamma_{z_0}^{+}$; call $s^1 = Rs^0$. Since J is slowly decreasing we can find real points (s^2, z^2) with

$$(7) \qquad |s^2 - s^1| + |z^2 - z^0| < a \log(1 + |s^1| + |z^0|)$$

and such that J is large on the set of all (s^2, z^2) thus obtained. We now use the minimum modulus theorem (see e.g. [5]): Fix z^2; we can find a circle $\delta(s^0, z^2)$ in the complex s plane with center s^1 and radius between

$$(8) \qquad 2n(a+c)\log(1 + |s^0| + |z^0|) \text{ and } 4n(a+c)\log(1 + |s^0| + z^0)$$

such that J is still large on the union of the sets $(\delta(s^0, z^2), z^2)$. Let s^3 be the point on $\delta(s^0, z^2)$ with largest imaginary value, that is, $Rs^3 = s^1$.

For $s^3, z_1^2, \cdots, z_{n-1}^2$ fixed, we can, by the above minimum modulus theorem, draw a circle ϵ in the complex z_n plane with center at z_n^2 and radius between

$$(9) \qquad a \log(1 + |s^1| + |z^0|) \text{ and } 2a \log(1 + |s^1| + |z^0|)$$

such that J is still large on the set of all points $(s^3, z_1^2, \cdots, z_{n-1}^2, \epsilon)$. Now, by construction, J has no zeros in ϵ. Moreover, z_n^0 lies in ϵ. Thus by the maximum modulus theorem J is large at the points $(s^3, z_1^2, \cdots, z_{n-1}^2, z_n^0)$.

Proceeding in this manner we find that J is large on the set of points $(s^3, z_1^0, \cdots, z_n^0)$. We can now "descend" from s^3 to s^0 by using the above minimum-maximum modulus argument on a chain of circles starting from s^3 with radii satisfying (9), noting that there will be no zeros in these circles.

This proves the Lemma for Γ_z^+ and the result for Γ_z^- is similar.

The same method leads also to

LEMMA 2. *There is an a' depending only on a, n, and the exponential type of J so that for any s above Γ_z^+ or below Γ_z^- we have*

$$(10) \qquad |J(s,z)| \geqq (a')^{-1}(1+|s|+|z|)^{-a'}\exp(-a'|Is|).$$

LEMMA 3. *The integrals in* (5) *converge in the topology of the space of distributions. Moreover, $H^\pm$ are fundamental solutions for S, that is, $S*H^\pm = \delta$.*

Proof. Denote by D the Fourier transform of the space D of indefinitely differentiable functions of compact support with the Schwartz topology. The topology of $\hat{D}$ is described in [4]: A fundamental system of neighborhoods of zero consists of those sets N_k of $F \in \hat{D}$ which satisfy

$$(11) \qquad |F(s,z)|/k(s,z) \leqq 1$$

where k is any continuous positive function on C such that any continuous function g on C for which there is a $d > 0$ so that

$$(12) \qquad g(s,z) = 0[\exp(d(|Is|+|Iz|)(1+|s|+|z|)^{-m}]$$

for all m must satisfy

$$(13) \qquad g(s,z) = 0(k(s,z)).$$

Now, any g which satisfies (12) will satisfy for $s \in \Gamma_z^\pm$

$$(14) \qquad g(s,z) = 0(1+|s|+|z|)^{-m}$$

for all m because z is real and because of (4). Using Lemma 1 the convergence of (5) follows easily.

It is clear that $S*H^\pm = \delta$.

It is also possible to prove that $H^\pm$ is a distribution by showing that $f \to H^\pm \cdot f$ defined by (5′) are continuous on every D_l where D_l is the space of C^∞ functions whose supports lie in the cube center origin side $2l$.

LEMMA 4. *The support of $H^\pm$ is contained in the set*

$$(15) \qquad t \leqq a' \qquad (\text{resp. } t \geqq -a').$$

Proof. We shall prove the result for H^+ as the result for H^- is similar. Let $f \in D$ and suppose that the support of f is in $t > a'$; call F the Fourier transform of f. Then we have, for $Is > 0$, z real,

$$\begin{aligned} |F(s,z)| &= |\textstyle\int f(t,x)\exp(its+ix\cdot z)\,dtdx| \\ &\leqq \textstyle\int |f(t,x)|\exp(-(a'+\epsilon)Is)\,dtdx \\ &= \exp(-(a'+\epsilon)Is)\textstyle\int |f(t,x)|\,dtdx, \end{aligned}$$

for some $\epsilon > 0$.

Applying this to the derivatives of f we deduce

$$|F(s,z)| = O((1+|s|+|z|)^{-m}\exp(-(a'+\epsilon)Is)) \tag{16}$$

for any $m > 0$.

Now,

$$H^{+}\cdot f = \int dz \int_{\Gamma_z^{+}} F(s,z)/J(s,z)\,ds.$$

For fixed z we may shift the contour Γ_z^{+} to $Is = +\infty$ using (16) and (10). It follows that $H^{+}\cdot f = 0$ which is the desired result.

Actually, Lemmas 3 and 4 are not sufficient for our purposes. To derive a stronger result we shall have to shift the contour in the integrals defining $H^{\pm}$ into the *complex* z space. For any complex z we define $\Gamma_z^{\pm}$ as the set of s satisfying

$$Is = \pm 4(a+c+1)(1+|Iz| + \log(1+|z|+|s|). \tag{17}$$

By the method of proof of Lemma 2 we have

LEMMA 5. *There is an a' depending only on a, n, and the exponential type of J so that for any s which is above Γ_z^{+} or below Γ_z^{-} we have*

$$|J(s,z)| \geqq (a')^{-1}(1+|s|+|z|)^{-a'}\exp(-a'(|Is|+|Iz|). \tag{18}$$

The method of proof of Lemma 3 leads to

LEMMA 6. *For any real numbers $u = (u_1,\cdots,u_n)$*

$$H^{\pm}(t,x) = \int e^{-ix\cdot z}\,dz \int_{\Gamma_z^{+}} e^{-its}\,ds/J(s,z) \tag{19}$$

where the z integration is over $Iz = u$. The integral in (19) *converges in the topology of the space of distributions.*

LEMMA 7. *There is an $a'' > 0$ so that the support of $H^{\pm}$ is contained in the set*

$$t \leqq a' - |x|/a'' \qquad (\text{resp.}\ t \geqq -a' + |x|/a''). \tag{20}$$

Proof. We shall consider H^{+} as H^{-} is treated by the same method. Suppose first that $n = 1$. Then we wish to show that H^{+} vanishes above the lines $t = a' - x/a''$ and $t = a' + x/a''$. Suppose that $f \in D$ and that the support of f is contained in $t > a' - x/a''$. Then if F denotes the Fourier transform of f we have, for $s \in \Gamma_z^{+}$, $Iz > 0$

$$
\begin{aligned}
|F(s,z)| &= \left|\int f(t,x)\exp(its+ixz)\,dtdx\right| \\
&\leqq \int |f(t,x)|\exp(-tIs-xIz)\,dtdx \\
&\leqq \int |f(t,x)|\exp(-4(a+c+1)t(1+z) \\
&\qquad -xIz)\,(1+|z|+|s|)^{-4(a+c+1)t}dtdx \qquad (21)\\
&= \int |f(t,x)|\exp(-4(a+c+1)t \\
&\qquad -(4(a+c+1)t+x)\,Iz)\,(1+|z|+|s|)^{-4(a+c+1)t}dtdx
\end{aligned}
$$

where we have used (17). Now, t stays in a bounded set so we have

$$|F(s,z)| \leqq e^{a_1}(1+|z|+|s|)^{a_1}\int |f(t,x)|\exp-(4(a+c)t+x)Iz\,dtdx$$

where $a_1 = \max 4(a+c+1)|t|$. Choose $a'' = 4(a+c)+1$ so

$$
\begin{aligned}
4(a+c+1)t+x &= a''t+x \qquad (22)\\
&\geqq a''a'+\epsilon
\end{aligned}
$$

for some $\epsilon > 0$ on the support of f. Thus

$$|F(s,z)| \leqq e^{a_1}(1+|z|+|s|)^{a_1}e^{-(a''a'+\epsilon)Iz}\int |f(t,x)|\,dtdx. \qquad (23)$$

Applying the same method to the derivatives of f we derive

$$|F(s,z)| \leqq a_2(1+|z|+|s|)^{-a'+4}e^{-(a''a'+\epsilon)Iz}. \qquad (24)$$

If we combine this with (18) and Lemma 6 then by letting $Iz \to +\infty$ we find $H^+ \cdot f = 0$. This shows that H^+ vanishes above $t = a' - x/a''$. The same method shows that t vanishes above $a' + x/a''$. This completes the proof in case $n = 1$.

The proof for $n > 1$ proceeds in a similar manner except that the lines $t = a' \pm x/a''$ are replaced by suitable affine hyperplanes.

3. Proof of Theorem 1. We assume first that the zeroes of J satisfy (1). Let $b > b_0 + a'$. Let $f \in E_b$ satisfy $S * f = 0$ on $|t| < b - b_0$. For any $\epsilon > 0$ there is a $g_\epsilon \in E$ which equals f on $|t| < b - \epsilon$ and $g_\epsilon = 0$ for $|t| > b + \epsilon$. We compute $S * g_\epsilon$:

$$S * g_\epsilon = h_\epsilon^1 + h_\epsilon^2 \qquad (25)$$

where support $h_\epsilon^1 \subset |t - b| < b_0 + \epsilon$ and support $h_\epsilon^2 \subset |t + b| < b_0 + \epsilon$.

Using Lemma 7 and standard methods we can prove

LEMMA 8. *The convolutions* $k_\epsilon^1 = H^- * h_\epsilon^1$ *and* $k_\epsilon^2 = H^+ * h_\epsilon^2$ *are in* E. *We have* $S * k_\epsilon^1 = h_\epsilon^1$ *and* $S * k_\epsilon^2 = h_\epsilon^2$. *Moreover,* $k_\epsilon^1 = 0$ *for* $t \leqq b - b_0 - \epsilon - a$ *and* $k_\epsilon^2 = 0$ *for* $t \geqq -b + b_0 + \epsilon + a'$.

We set

$$l_\epsilon = g_\epsilon - k_\epsilon^1 - k_\epsilon^2. \tag{26}$$

Then by Lemma 8 and (25), $S * l_\epsilon = 0$ and $l_\epsilon = f$ on $|t| < b - \epsilon - b_0 - a'$. It is easily seen that $f \rightarrow l_\epsilon$ is continuous from the topology of E_b into the topology of E.

Finally we must show that the extension is unique. Let $f \in E_{b'}$ satisfy $S * f = 0$ on $|t| < b_0$. Suppose $f = 0$ on $|t| < b$ where $b > b$. Then because the support of S is contained in $|t| < b_0$ we may write $f = f^+ + f^-$ where $f^+ = 0$, for $t < b$ and $f^- = 0$ for $t > -b$ and $S * f^+ = S * f^- = 0$. It remains to show that $S * f^+ = 0$. But

$$\begin{aligned} f^+ &= f^+ * \delta \\ &= f^+ * (S * H^+) \\ &= (f^+ * S) * H^+ \\ &= 0 \end{aligned}$$

where the use of the associativity of convolution is easily justified by use of Lemma 7. This completes the proof of the sufficiency of (1) for hyperbolicity.

We wish now to prove the necessity of (1). This can be done in two ways. The first is constructive: We use lacunary series of exponentials in the same manner as we used in [4] to prove the necessity of the conditions for ellipticity. We shall omit the details.

A second proof is by means of the closed graph theorem. Let b, b_1 be as in the definition of hyperbolicity. For any $f \in E_b$ such that $S * f = 0$ on $|t| < b - b_0$, we define $\delta_{\pm 2b} f$ by

$$\delta_{\pm 2b} f = g(\pm 2b, 0) \tag{27}$$

where $g \in E$ satisfies $S * g = 0$ and $g = f$ on $|t| < b - b_1$. Then by the definition of hyperbolicity, $\delta_{\pm 2b}$ are continuous on $E_b(S)$ which is the subspace of E_b consisting of those f which satisfy $S * f = 0$ on $|t| < b - b_0$. This means that there exists a neighborhood of zero in $E_b(S)$ on which $\delta_{\pm 2b}$ are bounded. Thus there are $k, d > 0$ so that if $h \in E$ satisfies $S * h = 0$ then

$$|h(\pm 2b, 0)| \leqq d \max_{|x| \leqq d,\, |t| \leqq b,\, \partial} |\partial h(t, x)| \tag{28}$$

where ∂ is any partial differential operator of the form $\partial / \partial x_1^{i_1} \cdots \partial x_n^{i_n}$ where $i \leqq k$.

We apply (28) to the functions $h(t, x) = \exp(its + ix \cdot z)$ with $(s, z) \in V$. Then we must have

$$\exp(2b\,|Is|) \leqq d'(1 + |s| + |z|)^k \exp(b\,|Is| + d\,|Iz|) \tag{29}$$

or

$$\exp(b\,|Is|) \leqq d'(1 + |s| + |z|)^k \exp(d\,|Iz|). \tag{30}$$

(30) is clearly equivalent to (1). This completes the proof of Theorem 1.

The above methods show also

THEOREM 2. *Let S be invertible. A necessary and sufficient condition that S be hyperbolic in t is that there exist $a', a'' > 0$ and fundamental solutions $H^\pm$ for S such that Lemma 7 holds.*

COROLLARY. *There exists a finite domain of dependence for solutions of $S^*f = 0$. More precisely, given any $t_0 > b_0 + a'$ (resp. $t_0 < -b_0 - a'$) and any x_0, the value $f(t_0, x_0)$ depends only on the values of $f(t, x)$ and its derivatives on*

$$|t| < 2b_0 + a' + \epsilon, \quad |x - x_0| < a''t + a'a'' \qquad (\text{resp. } |x - x_0| > -a''t - a'a'') \tag{31}$$

where ϵ is an arbitrarily small number.

Proof. Let $\alpha_\epsilon \in E$, $\alpha_\epsilon = 1$ on $|t| \leqq 2b_0 + a' + \epsilon/2$, $\alpha_\epsilon = 0$ on $|t| \geqq 2b_0 + a' + 3\epsilon/4$. Call $g_\epsilon = \alpha_\epsilon f$, and write $S^* g_\epsilon f = h_\epsilon^1 + h_\epsilon^2$ where

$$\text{support } h_\epsilon^1 \subset |t - 2b_0 - a' - \epsilon/2| < b_0 + \epsilon/4$$
$$\text{and support } h_\epsilon^2 \subset |t + 2b_0 + a' + \epsilon/2| < b_0 + \epsilon/4.$$

Call $k_\epsilon^1 = H^- * h_\epsilon^1$ and $k_\epsilon^2 = H^+ * h_\epsilon^2$ and set $l_\epsilon = g_\epsilon - k_\epsilon^1 - k_\epsilon^2$. Then by Lemma 8 $S^* l_\epsilon = 0$ and $l_\epsilon = f$ on $|t| < b_0 + \epsilon/4$. By uniqueness we have $l_\epsilon = f$, that is,

$$f = \alpha_\epsilon f - H^- * h_\epsilon^1 - H^+ * h_\epsilon^2. \tag{32}$$

Using Lemma 7 and the definition of h_ϵ^1, h_ϵ^2 we have the result.

THEOREM 3. *Let $n = 0$ and let S be a differential difference operator. Then S is hyperbolic.*

Proof. If S is the composition of a translation with an ordinary differential operator then J has only a finite number of zeros and the result is obvious. If S is not such a composition then the proof is the same as the proof of Lemma 5.18 of [4].

The analogous result for $n > 0$ is false, even if we assume S contains differences in t in a non-trivial manner as the example $n = 1$, $J = z\exp(is)$

shows. Even if $J(s,z)$ has the property that for no z does $J(s,z)$ vanish identically then the analogous result still fails as is shown by $n=1$, $J(s,z) = \cos z + \cos \lambda z \exp(is)$ where λ has suitable diophantine properties. As this example shows, in order for the partial differential-difference operator S to be hyperbolic in t we need suitable diophantine conditions on the translates in x which occur. If, however, S has no differences in any x_j and contains differences in t in a non-trivial manner, and if for no z does $J(s,z)$ vanish identically in s then a slight modification of the proof of Theorem 3 shows that S is hyperbolic in t.

We give now an example to show that the hypothesis "S is invertible" cannot be eliminated from Theorem 1. Let $n=0$ and set

$$J(s) = \prod_{j>1} [\cos s/j!j\log^2 j]^{j!}.$$

The zeros of J are real, but we shall show that J is not hyperbolic.

Call $l_j = j!j\log^2 j$ and set

$$f_j(t) = t^{j!-1}\exp(itl_j), \tag{33}$$

so $S^*f_j = 0$. We use the closed graph theorem as in the proof of Theorem 1. If S were hyperbolic then there would be numbers $0 < b < b'$ and $k, \alpha > 0$ so that for any solution f of $S^*f = 0$ we would have

$$|f(b')| \leqq \alpha \max_{|t|\leqq b} \sum_{i\leqq k} |f^{(i)}(x)|. \tag{34}$$

I claim that (34) cannot hold for all $f = f_j$ for any fixed α, k. For, the left side of (34) is

$$|f_j(b')| = (b')^{j!-1}.$$

Moreover, for $|t| \leqq b$ we have

$$\begin{aligned} |f_j^{(i)}(t)| &\leqq \sum_{r\leqq i} \binom{i}{r} (j!-1)^r |t|^{j!-1}(l_j)^{i-r} \\ &\leqq (k+1)!((j!)^{2k}b^{j!}l_j^k. \end{aligned}$$

In order for (34) to hold we would need

$$(b')^{j!} \leqq \alpha'(k+2)!(j!)^{2k}b^{j!}j^k(\log j)^{2k}. \tag{35}$$

It is clear that (35) cannot hold for all j since $b' > b$.

4. The generalized Cauchy problem for hyperbolic partial differential operators. In this section we shall assume that S is a partial differential operator of degree m which is hyperbolic in t. We call J the Fourier transform of S. Let $r \leqq m$, let $t_1 < t_2 < \cdots < t_r$, and let $\{\partial_{jk}\}_{\substack{j=1,\cdots,a_k \\ k=1,\cdots,r}}$ be linear partial differential operators with constant coefficients, where $\sum a_k = m$. Denote by ${}_0E$ the space of C^∞ functions of x with the Schwartz topology; ${}_0E'$ is the dual of ${}_0E$ and ${}_0\hat{E}'$ is its Fourier transform. We define the map $\alpha\colon E \to ({}_0E^m, E)$ where ${}_0E^m$ is the direct sum of m copies of ${}_0E$ by

$$(36) \qquad \alpha f = (\partial_{11} f(t_1, x), \cdots, \partial_{a_1} f(t_1, x), \cdots, \partial_{a_r} f(t_r, x), S * f).$$

Call α' the adjoint of α, so $\alpha'\colon (({}_0E^m)', E') \to E'$ is defined by

$$(37) \quad \alpha'(T_1, \cdots, T_m, W) \cdot f = \partial_{11} T_1 \cdot f(t_1, x) + \cdots + \partial_{a_r} T_m \cdot f(t_r, x) + S^* W \cdot f$$

for $f \in E$, $T_l \in {}_0E'$, $W \in E'$. The Fourier transform $\hat{\alpha}'\colon (({}_0\hat{E}')^m, \hat{E}') \to \hat{E}^0$ is therefore

$$(38) \quad \begin{aligned} &\hat{\alpha}'(F_1(z), \cdots, F_m(z), G(s, z)) \\ &\quad = P_{11}(s, z) F_1(z) \exp(it_1 s) + \cdots + P_{a_r r}(s, z) F_m(z) \exp(it_r s) \\ &\qquad\qquad + J(s, z) G(s, z) \\ &\qquad\qquad = H(s, z) \quad \text{say} \end{aligned}$$

where the P_{jk} are polynomials.

We wish to solve for F_j, G in terms of H. For fixed z let $s^l(z)$, $l = 1, \cdots, m$ be zeros of $J(s, z) = 0$. Suppose first, for simplicity, that the $s^l(z)$ are distinct for almost all z. Then for almost all z we have m equations

$$(39) \quad P_{11}(s^l, z) F_1(z) \exp(it_1 s^l) + \cdots + P_{a_r r}(s^l, z) F_m(z) \exp(it_r s^l) = H(s^l, z)$$

$$(40) \quad \begin{aligned} F_i(z) &= \Delta_i(z)/\Delta_0(z) \\ &= \Delta_i(z)\Delta_0(z)/[\Delta_0(z)]^2 \end{aligned}$$

where

$$(41) \qquad \Delta_0(z) = \det(P_{jk}(s^l, z) \exp(it_k s^l))$$

and a similar expression holds for $\Delta_i(z)$. We shall sometimes write $\Delta_i(H, z)$ for $\Delta_i(z)$.

In case $P(s, z)$ does not have distinct roots for almost all z, we differentiate (39) with respect to s sufficiently often and we derive m equations for the m unknowns $F_i(z)$ which are similar to (39) which can again be solved by Cramer's rule for almost all z. We shall refer to (40), (41) even for the general case.

Lemma 9. $[\Delta_0(z)]^2$ *and* $[\Delta_i(z)/\Delta_0(z)] \in {}_0\hat{E}'$. *Moreover,* $H \to \Delta_i(H, z)\Delta_0(z)$ *is continuous from* $\hat{E}'$ *into* ${}_0\hat{E}'$.

Proof. Consider $\Delta_0(z)$ first. We consider z and the s^l as independent variables so by (41) Δ_0 is an entire function of these variables. Moreover, it is easily seen that $[\Delta_0]^2$ is invariant under those permutations of the s^l which permute the roots of the irreducible factors of $J(s,z)$. Hence, if we write $[\Delta_0]^2$ as a power series, each homogeneous part is invariant under these permutations. By Galois theory each homogeneous part is a polynomial in z. This means we can write $[\Delta_0]^2$ as a power series in z. If we use the estimate $|s^l(z)| \leqq \text{const.}(1+|z|)$ which is a simple consequence of the fact that $t=0$ is not a characteristic for z, we find that this power series converges uniformly for z in compact sets. Hence $[\Delta_0]^2$ is entire.

If we now use (41) and (1) we see that $[\Delta_0]^2 \in {}_0\hat{E}'$.

The proof that $\Delta_l\Delta_0 \in {}_0\hat{E}'$ and the continuity of $H \to \Delta_l(H)\Delta_0$ is similar.

We want to prove that α is onto or, what can be shown by standard arguments to be the same thing, that α' is a closed mapping. Using [4] and (40) it would be sufficient to prove that $[\Delta_0]^2$ is slowly decreasing. Actually, however, it may be possible for $\Delta_0(z) \equiv 0$:

Example 1. $n=0$, $t_1=0$, $t_2=2\pi$, $\partial_1=\partial_2=$ identity, $S=\dfrac{d^2}{dx}+1$. Then $\Delta_0(z)\equiv 0$.

Example 2. $n=1$, $t_1=0$, $t_2=1$, $\partial_1=\partial_2=$ identity, $J=s^2+asz+bz^2+cs+dz+e$. Then $\Delta_0(z) \quad e^{is^1(z)}-e^{is^2(z)}$. Thus $\Delta_0(z)\equiv 0$ if, for example, $s^1(z)-s^2(z)=2\pi$. A simple computation shows that this is the case if and only if $b=a^2/4$, $d=ac/2$, $e=-\pi^2$.

It can be shown that α is never onto if $\Delta_0(z)\equiv 0$.

However, we have

THEOREM 4. *If* $\Delta_0(z)\not\equiv 0$ *then* $[\Delta_0(z)]^2$ *is slowly decreasing.*

For the proof we need

LEMMA 10. *Let* $n=1$. *Then the Puisseau expansion for* $s^l(z)$ *at infinity takes the form*

$$s^l(z)=c_1{}^l z+c_0{}^l+c_{-1}{}^l w^{-1}+\cdots \tag{42}$$

where $c_1{}^l$ *is real; here* $w=z^{1/m}$.

Proof. It is easily seen that w is a uniformizing parameter at infinity. We must show that if

$$s(z)=c_1z+c_0+e_1z^{1/m}+e_2z^{2/m}+\cdots+e_{m-1}z^{(m-1)/m} \tag{43}$$

then $s(z)$ can satisfy (1) only if c_1 is real and each $e_j=0$. From (43) we have

$$Is(z) = |c_1 z| \sin(\arg c_1 + \arg z) + Ic_0 + |e_1 z^{1/m}| \sin(\arg e_1 + 1/m \arg z) + \cdots$$
$$+ |e_{m-1} z^{(m-1)/m}| \sin(\arg e_{m-1} + (m-1)/m \arg z).$$

Letting $\arg z = 0$ we find that c_1 is real. Letting $\arg z = \pi$ we find $\arg e_{m-1} + (m-1)\pi/m = 0$, π, or 2π if $e_{m-1} \neq 0$. Letting $\arg z = 0$ we find $\arg e_{m-1} = 0$, or π. Thus $e_{m-1} = 0$; similarly all $e_j = 0$ which is the desired result.

Proof of Theorem 4. Assume first that $n = 1$. In (41) we replace each $s^l(z)$ by the expression (42). We have

$$\begin{aligned} \exp(is^l(z)t_k) &= \exp(it_k(c_1{}^l z + c_0{}^l + c_{-1}{}^l w^{-1} + \cdots)) \\ &= [\exp(it_k(c_1{}^l z + c_0{}^l))](1 + d_{-1}{}^l w^{-1} + \cdots) \end{aligned} \tag{44}$$

where the series converge in a neighborhood of $z = \infty$. Replacing the s^l in $P_{jk}(s^l, z)$ by (42) we deduce

$$[\Delta_0(z)]^2 = h_0(z) + h_{-1}(z) w^{-1} + \cdots \tag{45}$$

where the $h_p(z)$ are sums of pure imaginary exponentials and where the series converges in a neighborhood of $z = \infty$.

The above shows that not every $h_{-p}(z)$ can vanish identically. In fact we can show easily that for z real and sufficiently large, if p denotes the smallest subscript for which $h_{-p} \not\equiv 0$ then

$$[\Delta_0(z)]^2 = h_{-p}(z) w^{-p} + 0(w^{-p-1}). \tag{46}$$

Now h_{-p} is a finite sum of pure imaginary exponentials and so is almost periodic. Thus there exists an $\epsilon > 0$ and an L so that every interval of length L contains a point at which $|h_{-p}| > \epsilon$. Thus by (46) $[\Delta_0(z)]^2$ is slowly decreasing which is the desired result in case $n = 1$.

Passage to $n > 1$. Let $\lambda = (\lambda_1, \cdots, \lambda_n)$ be a parameter in C^n and denote by $[\Delta_0(z)]^2{}_\lambda$ the restriction of $[\Delta_0(z)]^2$ to the complex plane A_λ: $z_j = \mu\lambda_j$, where μ is a complex parameter. Call $\Delta_0{}^\lambda(\mu)$ the restriction of Δ_0 to A_λ. Then (46) becomes

$$[\Delta_0{}^\lambda(\mu)]^2 = h_{-p}{}^\lambda(\mu) \nu^{-p} + h_{-p-1}{}^\lambda(\mu) \nu^{-p-1} + \cdots \tag{47}$$

for μ real, $\mu \to \infty$ where $\nu = \mu^{1/m}$. We have

$$h_{-q}{}^\lambda(\mu) = \sum_a \beta_{qa}(\lambda) \exp(i\gamma_{qa}(\lambda)\mu) \tag{48}$$

where, as is easily seen, $\beta_{qa}(\lambda)$, and $\gamma_{qa}(\lambda)$ are integral algebraic functions of λ. Moreover $\gamma_{qa}(\lambda)$ is real for real λ.

We need the following Lemma.

LEMMA 11. *Let d be a positive integer. There exist positive constants η, ζ, ξ so that for any λ^0 and any $\epsilon > 0$ there is a real λ' with*

$$|\lambda' - \lambda^0| \leqq \epsilon \text{ and } |\beta_p(\lambda')| \geqq \eta\epsilon^\zeta. \tag{49}$$

We may choose $\lambda'(\lambda^0)$ in such a way that for all real λ^0 with $\sum |\lambda_j^0|^2 = 1$ we have for e_{qa} integers, $|e_{qa}| \leqq d$, oi

$$\min_{\substack{p \leqq q \leqq p+d \\ m' \text{ integer}}} \left|\sum e_{qa}\gamma_{qa}(\lambda') - m'\pi\right| \begin{cases} = 0 \\ or \geqq \xi \end{cases} \tag{50}$$

and

$$[\Delta_0^{\lambda'}(\mu)]^2 = h_{-p}^{\lambda'}(\mu)\nu^{-p} + \cdots + h_{-p-d}^{\lambda'}(\mu)\nu^{-p-d} + 0(\nu^{-p-d-1}) \tag{51}$$

for μ real where $0(\nu^{-p-d-1})$ is uniform in λ', ϵ for $\epsilon \leqq 1$, λ^0 real, $\sum |\lambda_j^0 = 1$. Moreover, for these values of λ' we have also a uniform bound for $|\beta_{qa}(\lambda')|$ and $|\gamma_{qa}(\lambda')|$ for all a and $p \leqq q \leqq p + d$.

Proof of Lemma 11. Let $\beta^1(\lambda) \not\equiv 0$ be an integral algebraic function and call $\beta^2, \cdots, \beta^s$ the conjugates of β. Then $\beta(\lambda) = \beta^1(\lambda)\beta^2(\lambda) \cdots \beta^s(\lambda)$ is a non zero polynomial. By performing a real linear transformation we may assume that

$$\beta(\lambda) = \lambda_1^b + \delta_1(\lambda_2, \cdots, \lambda_n)\lambda_1^{b-1} + \cdots + \delta_n(\lambda_2, \cdots, \lambda_n) \tag{52}$$

where the δ_j are polynomials. Applying the method of the author [1] or Malgrange [10] we find for each $\lambda_1, \lambda_2, \cdots, \lambda_n$ a set of the form $\{(\lambda_1'', \lambda_2, \cdots, \lambda_n)\}$ where λ_1'' lies in an interval I of length $\epsilon/2b$, the points of I being of distance $\leqq \epsilon$ from λ_1, such that

$$|\beta(\lambda_1'', \lambda_2, \cdots, \lambda_n)| \geqq \eta'\epsilon^b \tag{53}$$

where η' which depends only on b.

Now, since β^j are integral algebraic functions, and since λ varies in a compact set, the β^j are uniformly bounded on the sets of $(\lambda_1'', \lambda_2, \cdots, \lambda_n)$ which occur. Thus we have from (53)

$$|\beta^1(\lambda_1'', \lambda_2, \cdots, \lambda_n)| \geqq \eta''\epsilon^b. \tag{54}$$

Although we do not need it here, the same result holds in case β^1 is not integral. There is a polynomial θ whose product with β^1 is integral. We make a change of coordinates so that both θ and β are in the form (52). We then derive the inequality (54) for $\theta\beta^1$ instead of β^1. We can now "divide" by θ and obtain the desired inequality for β^1. The first part of Lemma 11 comes from setting $\beta_{p1} = \beta^1$.

The fact that (50) can be satisfied follows from the above considerations together with the facts that the λ' for which β^1 is large are free to vary over large sets and that the λ we are considering form a compact set.

If we use again the fact that the set of λ we need is compact and that we have much choice for the points λ', the result (51) becomes apparent. Similarly we have the boundedness of the $|\beta_{qa}(\lambda')|$ and $|\gamma_{qa}(\lambda')|$. This completes the proof of Lemma 11.

We shall need also

LEMMA 12. *Let* $A(\mu) = k_l(\mu)\mu^l + k_{l-1}(\mu)\mu^{l-1} + \cdots + k_0(\mu)$ *where the* k_j *are sums of pure imaginary exponentials. The distinct exponentials that occur are separated by an amount* u *and the absolute values of the coefficients are majorized by* v. *Write*

$$k_p(\mu) = \sum c_j{}^p \exp(ie_j{}^p\mu). \tag{55}$$

Then there exists B, y, $w > 0$ *which depend only on* u, v, l, η, ζ, m *and the number of exponentials that occur so that if for some* j_0 *and some real* μ^0 *with* $|\mu^0| \geqq B$

$$|c_{j_0}{}^l| \geqq \eta |\mu^0|^{-\zeta m} \tag{56}$$

then there is a real point μ' *with* $|\mu' - \mu^0| < w$ *and*

$$|A(\mu')| \geqq y |\mu^0|^{l-5-3[\lambda m]} \tag{57}$$

if $l > 5 + 3[\zeta m]$.

Proof. We may assume for simplicity that $e_{j_0} = 0$. Call

$$\psi_1(\sigma) = (\sin 8\pi u\sigma)^{l+2}/\sigma^{l+2}.$$

For each distinct $e_j{}^p \neq 0$ there is a unique zero $f_j{}^p$ of $\psi_1(\sigma)$ which is closest to $e_j{}^p$, and $f_j{}^m \neq 0$. We now "move" the zeros $f_j{}^m$ to $e_j{}^m$, that is, we call

$$\psi_2(\sigma) = \psi_1(\sigma) \prod (\sigma - e_j{}^p)^{l+2}/(\sigma - f_j{}^p)^{l+2}.$$

Finally we call

$$\psi(\sigma) = \frac{1}{\psi_2(0)} \left[\psi_2(\sigma) - \frac{d\psi_2(0)}{d\sigma}\sigma - \cdots - \frac{1}{l!}\frac{d^l\psi_2(0)}{d\sigma^l}\sigma^l\right].$$

It is easily seen that $\psi \in L_1(-\infty, \infty)$ and that the L_1 norm of ψ can be bounded in terms of u, l, and the number of $e_j{}^p$. Call χ the inverse Fourier transform of ψ. The methods of [2] show that $(\chi * A)(\mu) = Q(\mu)$ is the sum of those terms in A containing no exponential factor, that is,

$$Q(\mu) = c_{j_0}{}^l\mu^l + c_{j_0}{}^{l-1}\mu^{l-1} + \cdots + c_{j_0}{}^0.$$

Now, the support of χ is contained in $[-8\pi u, 8\pi u]$ and the maximum of χ is bounded by the L_1 norm of ψ.

From this we find easily that every real μ interval of length $16\pi u$ contains a point μ'' at which

$$|A(\mu'')| \geqq (16\pi u)^{-1} \| \psi \|_{L_1}^{-1} | Q(\mu'') |. \tag{58}$$

The proof of Lemma 12 will be complete if we can show that there exist $w', y' > 0$ which are independent of μ^0 so that the interval $|\mu - \mu^0| < w'$ contains a subinterval of length $16\pi u$ on which $|Q(\mu)| \geqq y' |\mu^0|^{l-5-8[\zeta m]}$.

We write c^p for $c_{j_0}{}^p$. We have

$$Q(\mu) = c^l(\mu - \mu_1)(\mu - \mu_2) \cdots (\mu - \mu_l). \tag{59}$$

We now appeal to a result of Montel (see M. Marden, *The Geometry of the Zeros of a Polynomial in a Comple Variable*, New York, 1949, p. 113). This asserts that there are at least $l - 4 - 2[\zeta m]$ zeros of Q in the circle

$$|\mu| < 1 + (v\eta^{-1} |\mu^0|^{\zeta m})^{\frac{1}{2}(1+\zeta m)} < |\mu^0|^{\frac{1}{2}}$$

if $|\mu^0| \geqq B$.

Label the μ_j so that $\mu_1, \cdots, \mu_{4+2[\zeta m]}$ are the μ_j with largest absolute value. If B is large enough (depending on ζm and u) then the set

$$\Gamma\colon |\mu - \mu^0| < (16\pi u + 2)(5 + 2[\zeta m])$$

will not meet $|\mu| < 2|\mu^0|^{\frac{1}{2}} + 2$. Since Γ contains at most $4 + 2[\zeta m]$ points μ_j, there is an interval I' in Γ of length $16\pi u + 2$ which is free of μ_j, hence a subinterval I of length $16\pi u$ which is at a distance at least 1 from the set of μ_j. Call $w = (16\pi u + 2)(5 + 2[\zeta m])$.

If we assume further that $B - B^{\frac{1}{2}} - w > B/2$ then we can estimate $Q(\mu)$ for $\mu \in I$ from below by

$$\begin{aligned} |\mu - \mu_j| &\geqq 1 \text{ for } j \leqq 4 + 2[\zeta m] \\ |\mu - \mu_j| &\geqq \tfrac{1}{2}|\mu| \text{ for } j > 4 + 2[\zeta m]. \end{aligned} \tag{60}$$

Combining this with (59) gives

$$|Q(\mu)| \geqq \eta^{2-l+4+2[\lambda m]} \, |\mu|^{l-4-2[\zeta m]-\zeta m}$$

which is the desired result. This completes the proof of Lemma 12.

Proof of Theorem 4 continued. Let z^0 be real and fixed. There is a real λ^0 with $\sum |\lambda_j{}^0|^2 = 1$ such that $z^0 \in A_0$. We want to find a point z' and a $w' > 0$ such that $|z' - z^0| \leqq 1 + w'$ and $[\Delta_0(z')]^2$ is large. The points z'' satisfying $|z''-z'| \leqq 1$ lie on lines $A_{\lambda''}$ with real λ'' satisfying $|\lambda''-\lambda^0| \leqq \epsilon = 1/|z^0|$. Let $\lambda'(\lambda^0)$ be as in Lemma 11, and let z''' be the point on $A_{\lambda'}$ which is closest to z^0. Write $z''' = \lambda'\mu^0$ where μ is as usual a complex parameter. In Lemma 11 we set $d = m(6+3[\zeta m])$. Then we have from (51),

$$[\Delta_0{}^{\lambda'}(\mu)]^2 = \nu^{-p-d}[H(\mu,\nu) + 0(\nu^{-1})] \tag{61}$$

where $H(\mu,\nu) = h_{-p}{}^{\lambda'}(\mu)\nu^d + \cdots + h_{-p-d}{}^{\lambda'}(\mu)$. Let $H_1, \cdots, H_m$ be the conjugates of H, that is, $H_j(\mu,\nu) = H(\mu, \nu \exp(2\pi ij/m))$, and call

$$A(\mu) = H_1(\mu,\nu)H_2(\mu,\nu) \cdots H_m(\mu,\nu)$$

so A is an exponential polynomial which is a polynomial of degree d.

We apply Lemma 12 to A. Thus there exists a μ' with $|\mu'-\mu^0| < w$ and $|A(\mu')| \geqq y|\mu^0|^{d-5-3[\zeta m]}$. On the other hand, by Lemma 11 we can find a $y' > 0$ which is independent of λ' such that

$$|H_2(\mu',\nu')H_3(\mu',\nu') \cdots H_m(\mu',\nu')| \leqq y'|\mu^0|^{(m-1)d/m}$$

where $(\nu')^m = \mu'$. This gives

$$|H(\mu',\nu')| \geqq y(y')^{-1}|\mu^0|.$$

Combining this with (61) we have

$$|\Delta_0{}^{\lambda'}(\mu')|^2 \geqq y''(\nu^0)^{-p}$$

for $|\mu^0| \geqq B_1$ with B_1 independent of λ'. This proves that $[\Delta_0{}^{\lambda'}(\mu)]^2$ is slowly decreasing which completes the proof of Theorem 4.

By the remarks at the beginning of this section we have

THEOREM 5. *If* $\Delta_0(z) \not\equiv 0$ *then* α *is onto.*

We wish to compute the kernel of α. First we note that every $H \in \hat{E}'$ can be written uniquely in the form

$$H(s,z) = F^1(z) + F^2(z)s + \cdots + F^m(z)s^{m-1} + J(s,z)G(s,z) \tag{62}$$

where the $F^j \in {}_0\hat{E}'$, $G \in \hat{E}'$ depend continuously on H. (This is the same as saying that the usual Cauchy problem is well-posed and appears in [6].) In particular (62) gives

$$P_{jk}(s,z)\exp(it_k s) = \sum F_{jk}{}^i(z)s^{i-1} + J(s,z)G_{jk}(s,z) \tag{63}$$

By the method of proof of Theorem 4, $\det(F_p{}^i)$ is either identically zero or is slowly decreasing.

Denote by $f_p{}^i$ the Fourier transform of $F_p{}^i$. If $\alpha g = 0$ then $S^* g = 0$. Thus g is uniquely determined by its usual Cauchy data say $g_j(x) = \frac{\partial^{j+1} g}{\partial t^{j+1}}(0, x)$ for $j = 1, \cdots, m$.

THEOREM 6. *The kernel of α consists of those g satisfying $S * g = 0$ and*

$$\sum f_j{}^i * g_i = 0 \quad \textit{for } j = 1, 2, \cdots, m. \tag{64}$$

Proof. $\alpha g = 0$ is equivalent to $\alpha'(W, 0) \cdot g = 0$ whenever

$$W = (W_1, \cdots, W_m) \in E'^r.$$

By (38) and (63)

$$\begin{aligned} \alpha'(W, 0) &= \sum_{jk} P_{jk}(s, z) \exp(it_k s) \hat{W}_{jk}(z) \\ &= \sum_{ijk} F_{jk}{}^i(z) \hat{W}_{jk}(z) s^i + J(s, z) \sum_{jk} G_{jk}(s, z) \hat{W}_{jk}. \end{aligned}$$

Since $S^* g = 0$ we find

$$\begin{aligned} \alpha'(w, 0) \cdot g = \sum_{ij} f_j{}^i * w_j \cdot g \\ + \sum_j w_j \cdot \sum_i f_j{}^i * g. \end{aligned}$$

Since $W \in E'$ is arbitrary we see that $\alpha g = 0$ is equivalent to (64).

From Theorem 6 and the definitions we have

COROLLARY. *Let* $\det(F_p{}^i(z^0)) = 0$. *Then there are non-zero constants c_i so that $g_i = c_i \exp(z^0 \cdot x)$ is a solution of* (64) *such that for the corresponding g we have $\alpha g = 0$.*

The example $n = 0$, $t_1 = 0$, $t_2 = 1$, $\partial_1 = \partial_2 =$ identity, $S = \frac{d^2}{ds^2}$ has $\det(F_p{}^i) = \sqrt{-1}$ and so has no zeros. For $n = 1$, $t_1 = 0$, $t_2 = 1$, $\partial_1 = \partial_2 =$ identity, $S = \frac{\partial^2}{\partial s^2}$ we find also $\det(F_p{}^i) = \sqrt{-1}$ so that we have examples where α is one-one even with $r > 1$. However, these examples seem to be exceptional.

Problem. In the proof of Theorem 4 we have had to pass from $n = 1$ to $n > 1$. It would be of interest to know if every $F \in \hat{E}'$ which is slowly decreasing on every complex line is slowly decreasing.

Explicit Formulae. Suppose $\Delta_0 \not\equiv 0$. If $(f_1, \cdots, f_m, g) \in ({}_0E^m, E)$ then an f satisfying $\alpha f = (f_1, \cdots, f_m, g)$ is given as follows: Let $K_i(z; t^0, x^0) = \Delta_i(z) \Delta_0(z)$ where $\Delta_i(z)$ is evaluated for $\exp(it^0 s + ix^0 \cdot z)$. Then by (38) ff. we can write

$$\begin{aligned}\exp(it^0s+ix^0\cdot z)[\Delta_0(z)]^3=P_{11}(s,z)K_1(z;t^0,x^0)\exp(it_1s)+\cdots\\ +P_{a_rr}(s,z)K_m(z;t^0,x^0)\exp(it_rs)\\ +J(s,z)L(s,z;t^0,x^0).\end{aligned}$$

Call $T_j(t^0,x^0)$, $W(t^0,x^0)$ the inverse Fourier transforms of K_j and L. Then we can write

$$f(t^0,x^0)=\sum T_j(t^0,x^0)h_j+W(t^0,x^0)\cdot k$$

where h_j, k satisfy $U*h_j=f_j$, $U*k=g$, U being the inverse Fourier transform of Δ_0^2.

Remark. The Cauchy problem can be generalized still further. Namely, let $A_1,\cdots,A_m$ be any invertible elements of E'. We consider the map α: $E\to({}_0E^m,E)$ defined by $\alpha f=((A_1*f)(0,x),\cdots,(A_m*f)(0,x),S*f)$. The case treated above is when A_j is of the form $\partial\tau$ where τ is a translation in t and ∂ is a linear constant coefficient partial differential operator. In case the support of A_j is a finite set, then the above method extends easily. For $n=1$ we can give simple criteria so that the analog of Theorem 4 holds. However, the passage to $n>1$ seems to involve great difficulties.

5. The Cauchy problem for hyperbolic convolution equations and Wiener-Hopf problems. Let S be hyperbolic in t. We shall assume for simplicity that $-1\leqq t\leqq 1$ is the smallest t strip containing the support of S. By E_1 we denote the space of functions which are C^∞ for all x and for $-1\leqq t\leqq 1$. Let $a(D)$ be a linear function on the space of $D\in E'$ whose support is the hyperplane $t=0$. We define $E_1(S,a)$ as the subspace of E_1 consisting of those h which satisfy $D*\check{S}\cdot h=a(D)$, with $\check{S}(t,x)=S(-t,-x)$.

Cauchy Problem. Given $g\in E$ and $h\in E_1(S,a)$ find $f\in E$ such that $S*f=g$ and $f=h$ on $-1\leqq t\leqq 1$, where $a(D)=D\cdot g$.

(If S were a differential operator the interval $-1\leqq t\leqq 1$ would degenerate to the hyperplane $t=0$ and the above Cauchy problem would be the usual one.)

Denote by α the map of $E\to(E_1,E)$ defined by

$$\alpha f=(\text{restriction of } f \text{ to } -1\leqq t\leqq 1, S*f).$$

$\alpha'\colon(E_1',E')\to E'$ is defined by

$$\alpha'(T,W)=T+S*W$$

where $T\in E_1'$ is considered in a natural way as an element of E'. The Fourier transform $\hat{\alpha}'$ of α' is

$$\hat{\alpha}'(G,H)=G+JH. \tag{65}$$

It is clear that the image of α lies in the subspace L of $E_1 \oplus E$ consisting of those pairs (h, g) with $S \cdot Dh = D \cdot g$ for any D. We say that *the Cauchy Problem is well-posed* if $\alpha: E \rightarrow L$ is a topological isomorphism. The dual L' of L is the quotient of $E_1' \oplus E'$ by the space of (T, W) which satisfy $T \cdot h + W \cdot g = 0$ for $(h, g) \in L$. That is, there is a D such that $T = D\check{S}$, $W = -D$. By Fourier transform, $G = PJ$, $H = -P$ for some P which is a polynomial in s, where G, H are the Fourier transforms of T, W.

Now, let us go back to (65). Since $G \in \hat{E}_1'$, $G \in \hat{E}'$ and G is of exponential type $\leqq 1$ in s. It follows immediately from the theorem of addition of supports (see [9]) that $\hat{\alpha}'(G, H) = 0$ if and only if $G = PJ$, $H = -P$ for some P which is a polynomial in s. Thus α' is one-one on L' which means that the image of α is dense in L.

We have the much stronger

THEOREM 7. *If S is hyperbolic in t and invertible the Cauchy problem is well-posed.*

Proof. We want to represent any $K \in \hat{E}'$ in the form $K = \hat{\alpha}'(G, H)$ with $(G, H) \in \hat{L}'$ depending continuously on K. Let d be a positive integer depending on K which will be prescribed later. For fixed z we use the curves $\Gamma_z^{\pm}$ defined by (4) for $c' = 4(a + c)$ in the notation of Lemma 1. We set

$$G(s, z) = s^d J(s, z) \left[\int_{\Gamma_z^+} + \int_{\Gamma_z^-} \right] K(\zeta, z)\, d\zeta / \zeta^d J(\zeta, z)(s - \zeta) \tag{66}$$

if s does not lie between $\Gamma_z^{\pm}$ and

$$G(s, z) = s^d J(s, z) \left[\int_{\Gamma_z^+} + \int_{\Gamma_z^-} \right] K(\zeta, z)\, d\zeta / \zeta^d J(\zeta, z)(s - z) - K(s, z), \tag{67}$$

if s lies betwen $\Gamma_z^{\pm}$ where $\Gamma_z^{\pm}$ are oriented so that the interior lies on the left.

LEMMA 13. *$G(s, z) \in \hat{E}_1'$ and $K - G$ is divisible by J in $\hat{E}'$.*

$(G, (K - G)/J) \in L'$ depend continuously on $K \in \hat{E}'$.

Proof. We use the methods of [4] Section 3. Suppose for $a'', a''' > 0$,

$$|K(s, z)| \leqq \alpha(1 + |s| + |z|)^{a''} \exp(a'''(|Is| + |Iz|)). \tag{68}$$

then by Lemma 1, if we choose

$$d = 4 + [a' + a'' + 4(a + c)a'''] + d' \tag{69}$$

(where d' is any positive integer) the integrals (66) or (67) converge for s off the curves $\Gamma_z^{\pm}$.

Now, let $\{s^j\}^{\infty}_{j=-\infty}$ be the points of Γ_z^+ with real parts

$$s^j = 32j(a+c)(1+\log(1+|z|+|s^j|)).$$

Using s^j as center, we may by the minimum modulus theorem (see [5]) draw a circle γ_z^j of radius between $2(a+c)(1+\log(1+|z|+(s^j|))$ and $3(a+c)(1+\log(1+|z|+|s^j|))$ on which

$$|J(s,z)| \geqq \operatorname{const}(1+|s|+|z|)^{-b}$$

where b is independent of j and z. If we use the right arc of γ_z^j joining Γ_z^+ to Γ_z^- and if we let d' in (69) be $[b]+2$ then we can, by Cauchy's theorem write in (66) or (67)

$$\int_{\Gamma_z^+} + \int_{\Gamma_z^-} = \sum_j \int_{\Gamma_z^j} \tag{70}$$

where Γ_z^j is the contour bounded by $\Gamma_z^{\pm}$ and the arcs of γ_z^{j-1} and γ_z^j chosen. The series on the right of (70) converges absolutely for s of $\bigcup\limits_j \Gamma_z^j$.

It now follows immediately from the results of [4] Section 3 that for fixed z, $G(s,z)$ is entire in s.

Next, we notice that if z varies in a small set U then by Cauchy's theorem and the choice of the $\Gamma_z^{\pm}$ we have $\int_{\Gamma_{z^0}^{\pm}} = \int_{\Gamma_{z^0}^{\pm}}$ for any z^0 in U. It follows from (66) and (67) that $G(s,z)$ is for fixed s holomorphic for $z \in U$. Hence $G(s,z)$ is entire in s and z.

We can now estimate $G(s,z)$ from (66) and (67). Suppose first that distance $(s,\Gamma_z^{\pm}) \geqq 1$. Then using Lemma 1, (68) and (69) we have

$$\left| \int_{\Gamma_z^{+}} + \int_{\Gamma_z^-} \right| \leqq \operatorname{const}(1+|z|)^{a'+a''+4(a+c)a'''} \exp(a'''|z|). \tag{71}$$

Since the contours $\Gamma_z^{\pm}$ can be moved by shifting $s \to s+2i$ without changing anything except the constants, (71) holds for all (s,z). Hence using (66) and (67) we have

$$|G(s,z)| \leqq \beta(1+|s|)^d |J(s,z)|(1+|z|)^{\beta'} \exp(a'''|Iz|) \tag{72}$$

where β depends on J and a, a'', a''', and β' depends on J, a'', a''', and d depends on J, a'', and a'''.

From (72) it follows first that G is an entire function of exponential type in $\hat{E}_1'$. Using the dependence of β, d, β' on K it follows also that if K varies in a bounded set so does G.

We use again (70) and the results of [4] Section 3. It follows that for fixed z, $K - G$ is divisible by J in the ring of entire functions of s. A simple estimate shows that $(K-G)/J$ is bounded in the neighborhood of any point (s,z). As is well known, this implies that $(K-G)/J$ is entire. By Theorem 2.6 of [4] $(K-G)/J$ is in $\hat{E}'$.

By the above and Theorem 2.2 of [4], if K varies in a bounded set of $\hat{E}'$, then G and $(K-G)/J=H$ vary in bounded sets of $\hat{E}_1{}'$, $\hat{E}'$ respectively.

If we change d to another integer satisfying (69) then (G,H) is replaced by a pair (G_1,H_1) which is equal to (G,H) considered as an element of $\hat{L}'$ because $K=G+JH=G_1+JH_1$. Thus, $K\to(G,H)$ is a linear map of $\hat{E}'\to L'$ which is bounded on the bounded sets of $\hat{E}'$. Since $\hat{E}'$ is bornologic, this map is continuous. This completes the proof of Lemma 13.

Proof of Theorem 7 concluded. It is clear that α' is continuous. Moreover, the map of $\hat{E}'\to\hat{L}'$ defined above by $K\to(G,H)$ is, because $K=G+JH$, an inverse for $\hat{\alpha}'$. By Lemma 13 this inverse is continuous. This proves that $\hat{\alpha}'$ hence α is a topological isomorphism which completes the proof of Theorem 7.

Relation to Wiener-Hopf problems. Formally for $n=0$ the Wiener-Hopf problem can be stated as follows: Let k be a function. Then we want to find functions f' which satisfy

$$\int_0^\infty k(t-t')f'(t')\,dt'=f'(t) \text{ for } t>0. \tag{73}$$

If we define f to be f' for $t>0$ and 0 for $t\leqq 0$ then (73) is equivalent to

$$\begin{aligned}(k*f)(t)&=f(t) &&\text{for } t>0\\ f(t)&=0 &&\text{for } t\leqq 0.\end{aligned} \tag{74}$$

Next, let k' be the measure $k'=k-\delta$. Then (74) becomes

$$\begin{aligned}(k'*f)(t)&=0 &&\text{for } t>0\\ f(t)&=0 &&\text{for } t\leqq 0.\end{aligned} \tag{75}$$

Let us assume that k, hence k' is of compact support. Then we may generalize (75) by replacing k' by an arbitrary $S\in E'$. We also look for $f\in E$. Thus we want

$$\begin{aligned}(S*f)(t)&=0 &&\text{for } t>0\\ f(t)&=0 &&\text{for } t\leqq 0.\end{aligned} \tag{76}$$

Suppose $[a,b]$ is the smallest interval containing the support of S. Then to solve (76) it is sufficient to solve

$$(77)\qquad \begin{array}{ll} (S^{*}f)(t)=0 & \text{for } t>0 \\ f(t)=0 & \text{for } -b\leqq t\leqq 0 \end{array}$$

if $b\geqq 0$ while the second condition in (77) is vacuous if $b<0$.

Assume that $0\in[a,b]$. Then (77) is just a special case of the Cauchy problem (or rather the initial value problem which can be treated by the same methods)

$$(78)\qquad \begin{array}{ll} (S^{*}f)(t)=g(t) & \text{for } t>0 \\ f(t)=h(t) & \text{for } a-b\leqq t\leqq 0 \end{array}$$

namely we specialize to $g=0$ and $h=0$ for $-b\leqq t\leqq 0$.

If we now let $n>0$ then it is clear that the obvious extension of (77) is a special case of the Cauchy problem treated in this paper; here $a\leqq t\leqq b$ is the smallest strip containing the support of S.

We wish to show that, under a suitable hypothesis, the Wiener-Hopf problem has solutions for any n. Let 0 be an inner point of $[a,b]$. We assume for simplicity of notation that $a=-1$, $b=1$. By D we denote the space of Schwartz indefinitely differentiable functions of compact support. For any $l\in D$, let K be its Fourier transform. Then as in Lemma 13 we write $K=G+JH$.

From (72) and a check of the dependence of the constants, we have

$$(79)\qquad |G(s,z)|\leqq\beta\,|J(s,z)|(1+|z|)^{\beta'}\exp(\alpha\,|Iz|)$$

Since J is slowly decreasing it follows that the H for which $G+JH\in\hat{D}$ are polynomials in s *with bounded degree.*

Next, for $h\in D$, support $h\subset\frac{1}{4}\leqq t\leqq\frac{3}{4}$ we want to define $F\in D'$ by

$$(80)\qquad f\cdot l=h\cdot T$$

where T is the inverse Fourier transform of G.

From Lemma 13, f is bounded on the bounded sets of D so f is a distribution. Now,

$$\begin{aligned}(S^{*}f)\cdot l&=f\cdot S^{*}l\\&=h\cdot T\end{aligned}$$

where the Fourier transform of T is the G corresponding to JK. It is clear that $G=0$ so $S^{*}f=0$.

Let the support of l be contained in $[-1,1]$. Then $f\cdot l=h\cdot T$. Since $K=G+JH$, $T=l-S^{*}W$ where W is the inverse Fourier transform of H.

Thus, $f \cdot l = h \cdot l - h \cdot S^*W$ where by our above remarks, H is a polynomial in s of degree $\leqq m$.

Require. $h \cdot S^*W = 0$ for all W of the form

$$W = W_0 + \partial/\partial t W_1 + \cdots + \partial^m/\partial t^m W_m$$

where W_j is a distribution in x.

THEOREM 8. *Suppose there exists an h satisfying the above requirement. Then there exists a solution of* (77). *If the space of h satisfying the requirement is infinite dimensional then so is the space of solutions of* (77).

Proof. f is a distribution which satisfies $S^*f = 0$ and restriction of f to $-1 < t < 1$ equals h. If we convolve f with an element of D of small support then we have an element $f_1 \in E$ with $S^*f_1 = 0$, $f_1 = 0$ on $-\frac{7}{8} \leqq t \leqq \frac{1}{8}$, $f_1 \not\equiv 0$. We define f_2 as the translate of f_1 by an amount in $t = \frac{1}{8}$. Then $S^*f_2 = 0$, $f_2 = 0$ on $-1 \leqq t \leqq 0$, $f_2 \not\equiv 0$. The space of f_2 thus constructed is clearly infinite dimensional and we are finished.

Remark. It may not always be true that there exist h satisfying the requirement in the hypothesis of Theorem 8. For example, let

$$S \cdot g = \sum b_{j_1 \cdots j_{n-1}} \int_{-1}^{1} g(t, 1/j_1, \cdots, 1/j_n) \exp(it(j_1 + \cdots + j_n))\, dt \tag{81}$$

where $b_{j_1 \cdots j_n}$ are positive numbers whose sum is 1. We have

$$\begin{aligned} J(s,z) = 2 \sum b_{j_1 \cdots j_n} \exp(iz_1/j_1) \cdots \\ \times \exp(iz_n/j_n) \sin(s + j_1 + \cdots + j_n)/(s + j_1 + \cdots + j_n). \end{aligned} \tag{82}$$

It seems that for suitable $b_{j_1 \cdots j_n}$ we have J slowly decreasing and S hyperbolic in t. It seems also that the set of S^*W is dense in the space of distributions on $-\frac{3}{4} < t < \frac{3}{4}$ where W varies over the distributions in x. Thus no h could exist.

Of course, if $n = 0$ then there are only a finite number of conditions for h to satisfy so the space of h is infinite dimensional. If $n > 0$ and S is a finite sum of products of distributions in t with distributions in x, we can for the same reason find an infinite dimensional set of h of the form $h_1(t) h_2(x)$.

Remark. Let $n = 0$. The compatibility conditions for the Cauchy data $(h, g) \in L$ are infinite in number. Presumably for any m we could construct spaces L^m, E^m, which depend on S, such that the Cauchy problem is well posed for $f \in E^m$, the Cauchy data in L^m satisfying m prescribed conditions. It would be of interest to study these spaces. Similar remarks hold for $n > 0$.

Explicit Formulae. If $h, g \in L$ then the explicit form of f satisfying $\alpha f = (h, g)$ is

$$f(t, x) = h \cdot S_{t,x} + g \cdot W_{t,x} \tag{83}$$

where

$$\exp(its + ix \cdot z) = G_{t,x}(s, z) + J(s, z) H_{t,x}(s, z). \tag{84}$$

REFERENCES.

[1] L. Ehrenpreis, "Solutions of some problems of division I," *American Journal of Mathematics*, vol. 76 (1954), pp. 883-903.

[2] ———, "Solutions of some problems of division II," *ibid.*, vol. 77 (1955), pp. 286-292.

[3] ———, "Solutions of some problems of division III," *ibid.*, vol. 78 (1956), pp. 685-715.

[4] ———, "Solutions of some problems of division IV," *ibid.*, vol. 82 (1960), pp. 522-588.

[5] ———, "Mean periodic functions I," *ibid.*, vol. 77 (1955), pp. 293-328.

[6] ———, "Cauchy's problem for linear differential equations with constant coefficients," *Proceedings of the National Academy of Sciences*, vol. 42 (1956), pp. 642-646.

[7] ———, "A fundamental principle and some of its applications," Proceeding of the Symposium on Linear Spaces, Jerusalem, 1960.

[8] L. Gårding, "Linear hyperbolic equations with constant coefficients," *Acta Mathematica*, vol. 85 (1951), pp. 1-62.

[9] J. L. Lions, "Supports de produits de composition," *Comptes Rendus*, vol. 232 (1951), pp. 1530-1532.

[10] B. Malgrange, "Existence et approximation des solutions des équations aux dérivées partielles et des équations de convolution," thesis, Paris, 1955, Annals l'Intitut Fourier (Grenoble), vol. 6 (1956), pp. 271-355.

[11] R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex domain*, New York, 1934.

ON EXTENSIONS OF ALGEBRAS, CO-ALGEBRAS AND HOPF ALGEBRAS, I.*

By V. K. A. M. GUGENHEIM.[1]

The classical theory of extensions could be described briefly by saying that it is about deviations from the additive structure of the direct sum; the present paper is about deviations from the multiplicative structure of the tensor product.

One of the most striking features of the new theory is its thorough-going duality. The basic structure of the classical theory is an exact sequence $0 \to A \to C \to B \to 0$ and this is—in the sense of Buchsbaum, Grothendiek and others—a self-dual concept. Accordingly, the classical theory is basically self-dual and appears, essentially, as one.

The notion of an algebra is not self-dual; on the contrary there is the dual notion of a co-algebra; and thus the present theory appears in two dual aspects: a theory of extensions for algebras, and one for co-algebras. Not surprisingly, however, the most interesting situation is precisely that which is again self-dual: the extension theory for Hopf algebras.

In the present paper, duality has been stressed to the point of pedantry: Every statement n_* corresponds to a dual statement n^*; statements labelled n are self-dual; see also the remarks following 1.1^*.

Extensions $A \to C \to B$ are in many ways analogous to fibre bundles with fibre A, base-space B; this is apparent in the theory of "induced extensions"; in some sense, the induced fibre space exists because the diagonal map $B \to B \times B$ is commutative; and it will be seen that we get a good theory of induced structures only if the algebra A, or else the co-algebra B, is commutative.

In the theory of central extensions of Hopf algebras we have all these structures: they combine to define in this theory a group structure analogous to that introduced by Baer into the classical one.

The "extensions" of this paper occur very naturally in the theory of Hopf algebras: the basic lemmas 2.5, 2.8 in the paper [1] of Milnor and

---

* Received November 30, 1961.

[1] The author was partially supported by Grant DA-ORD-31-124-61-G60, Office of Ordnance Research.

Moore exhibit a Hopf algebra as such an extension; the computation of the cohomology of the Steenrod algebra due to Adams, [2], proceeds by exhibiting that algebra as a sequence of extensions.

Just as the classical theory, the present one is related to certain invariants of a cohomological kind. The theory of these invariants will be the subject of a subsequent paper; they are related to the twisting co-chains of E. H. Brown, [3]; for some first indications, see [4].

The paper falls, roughly, into three parts. Chapters 1-4 develop the necessary properties of the tensor product and co-tensor product; the former are "well-known"—but not in the form needed for their dualisation. Chapters 5, 6 give the theory of extensions. Chapter 7 is a technical digression on the "bi-tensor product." This subsumes both the dual products used earlier and is, in a certain sense, the underlying construction in the case of Hopf algebras. Chapter 8 gives an almost self-contained alternative theory of central extensions; it is less invariant, more explicit, and used here mainly to complete the proof of Theorem 6. 61.

Finally, a point of notation: if X is an object in some category, the same symbol X will be used to denote the identity map $X \to X$.

It is a pleasure to express my thanks to J. C. Moore for many essential suggestions both of a general and of a specific kind; in particular, he pointed out to me the existence of the co-tensor product, and its relevance to my—at that time somewhat visionary—investigation.

1. Preliminaries. Let K be a commutative ring with unit $1 \in K$. A sequence $A = \{A_n\}$ of K-modules will be called a graded module, a sequence $f = \{f_n \colon A_n \to B_n\}$ a map of graded modules. We shall often identify K with the graded module $\{K_n\}$ with $K_n = 0$, $n \neq 0$, $K_0 = K$.

The graded module A will be called a G. C. (graded connected) module if $A_n = 0$ for $n < 0$ and there are given maps of graded modules $\epsilon \colon A \to K$, $\eta \colon K \to A$ such that $\epsilon_0 \eta_0 = K$, $\eta_0 \epsilon_0 = A_0$. A map of graded modules must be consistent with this structure—i. e., "keep fixed" the elements of $A_0 = K$. If A, B are G. C. modules, the G. C. module $A \otimes B$ is defined by $(A \otimes B)_n = \sum_{p+q=n} A_p \otimes B_q$. We shall invariably identify A, $A \otimes K$, $K \otimes A$ under the isomorphisms $a \to a \otimes 1 \to 1 \otimes a$.

1. 1.$_*$ *Definition.* The pair (A, ϕ) will be called an algebra if

(i) A is a G. C. module

(ii) ϕ is a map of G. C. modules $A \otimes A \to A$

(iii) The following diagrams are commutative

$$\begin{tikzcd}
A \otimes A \otimes A \arrow[r, "A \otimes \phi"] \arrow[d, "\phi \otimes A"] & A \otimes A \arrow[d, "\phi"] \\
A \otimes A \arrow[r, "\phi"'] & A
\end{tikzcd}$$

$$\begin{tikzcd}
K \otimes A \arrow[r, "\eta \otimes A"] \arrow[dr, equal] & A \otimes A \arrow[d, "\phi"] \\
& A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \otimes K \arrow[r, "A \otimes \eta"] \arrow[dr, equal] & A \otimes A \arrow[d, "\phi"] \\
& A
\end{tikzcd}$$

1.1* *Definition.* The pair (B, ψ) will be called a co-algebra if

(i) B is a G. C. module

(ii) ψ is a map of G. C. modules $B \rightarrow B \otimes B$

(iii) The following diagrams are commutative

$$\begin{tikzcd}
B \otimes B \otimes B & B \otimes B \arrow[l, "B \otimes \psi"'] \\
B \otimes B \arrow[u, "\psi \otimes B"] & B \arrow[l, "\psi"] \arrow[u, "\psi"']
\end{tikzcd}$$

$$\begin{tikzcd}
K \otimes B \arrow[dr, equal] & B \otimes B \arrow[l, "\epsilon \otimes B"'] \\
& B \arrow[u, "\psi"']
\end{tikzcd}
\qquad
\begin{tikzcd}
B \otimes K \arrow[dr, equal] & B \otimes B \arrow[l, "B \otimes \epsilon"'] \\
& B \arrow[u, "\psi"']
\end{tikzcd}$$

The two above definitions are "dual" to each other; such dual statements will occur many times in the present paper. The statement dual to n_* will always be denoted by n^*; in order to save space often only one of n_*, n^* will be written out; nevertheless, the other one will in each case be implied, and referred to by its appropriate number even if it is nowhere written down.

1.2* *Definition.* Let (B, ψ) be a co-algebra. The pair (N, σ) will be called a left B-comodule if

(i) N is a G. C. module

(ii) σ is a map of G. C. modules $N \to B \otimes N$

(iii) The following diagrams are commutative

$$\begin{tikzcd}
B\otimes B\otimes N & B\otimes N \arrow[l, "B\otimes\sigma"'] \\
B\otimes N \arrow[u, "\psi\otimes N"'] & N \arrow[l, "\sigma"] \arrow[u, "\sigma"']
\end{tikzcd}
\qquad
\begin{tikzcd}
K\otimes N & B\otimes N \arrow[l, "\epsilon\otimes N"'] \\
 & N \arrow[u, "\sigma"'] \arrow[ul, equal]
\end{tikzcd}$$

An analogous definition is given for a right "B-comodule." The dual definition 1.2_* for a left A-module (M, τ) is the classical one with $\tau(a\otimes m) = am$.

1.3^* *Definition.* Let (B, ψ), (B', ψ') be coalgebras. A map of G. C. modules $g\colon B \to B'$ will be called a map of co-algebras if

$$(g\otimes g)\psi = (\psi'\otimes\psi')g.$$

1.4^* *Definition.* Let (N, σ) be a left B-comodule, (N', σ') a left B'-comodule and $f\colon B \to B'$ a map of G. C. modules. A map of G. C. modules $g\colon N \to N'$ will be called an f-map if

$$(f\otimes g)\sigma = \sigma' g.$$

Remark. If the composition

$$N \xrightarrow{\sigma} B\otimes M \xrightarrow{B\otimes\epsilon} B\otimes K = B$$

is an epimorphism, it follows that f is a map of co-algebras; this will certainly be true in most applications.

If, in 1.4^*, f is the identity, we call g a map of B-comodules.

Convention. From now on we shall usually refer to an algebra A, a co-algebra B, a left A-module M etc.; allowing the symbols ϕ, ψ, τ to be understood. In order to be able to do this easily we shall denote all "multiplications" by ϕ, all "co-multiplications" by ψ, all "operations" by τ, all "co-operations" by σ; and, in case of ambiguity, will write ϕ_A, ψ_B, τ_M, σ_N, etc. Also, extending a notation of Cartan-Eilenberg we shall denote by ${}_AM$; ${}^BN$ the "situations" of 1.2_*, 1.2^* respectively; and we shall use the shorthand $M = {}_AM$ for the phrase "M is a left A-module," etc.

1. 5_* *Definition.* We shall refer to the situation $M = {}_AM_{A'}$ if (M, τ) is a left A-module, (M, τ') a right A'-module and

$$\tau(A \otimes \tau') = \tau'(\tau \otimes A').$$

Note that $A = {}_AA_A$ and $B = {}^BB^B$.

2. Tensor products and co-tensor products. If A, B are G. C. modules and $\Delta\colon A \to B$ a map of graded modules such that $\Delta_0 = 0$, then kernel Δ and co-kernel Δ are G. C. modules.

2.1^* *Definition.* Let $M = M^B$, $N = {}^BN$ where B is a co-algebra. We define

$$\Delta = \Delta^B = \sigma_M \otimes N - M \otimes \sigma_N\colon\ M \otimes N \to M \otimes B \otimes N$$

and $M \otimes {}^BN = \ker \Delta^B$; cf. [5], Exp. 7, p. 25.

We refer to $M \otimes {}^BN$ as the co-tensor product of M and N over B. The dual concept, denoted by $M \otimes {}_AN$, is the well known product of two A-modules; we shall use the notation $\Delta = \Delta_A = \tau_M \otimes N - M \otimes \tau_N$ so that $M \otimes {}_AN = \text{co-ker}\, A$.

We shall say that the sequence $0 \to A \xrightarrow{i} B \xrightarrow{j} C$ is split exact if it is exact and image i, image j are direct summands; $A \xrightarrow{i} B \xrightarrow{j} C \to 0$ is split exact if it is exact and $\ker i$, $\ker j$ are direct summands.

2.2^* Lemma. *Let $N = {}^BN$. The sequence*

$$0 \to N \to B \otimes N \xrightarrow{\Delta} B \otimes B \otimes N$$

where $\Delta = \psi \otimes N - B \otimes \sigma$, is split exact.

This lemma says $N = B \otimes {}^BN$. The dual Lemma 2.2_* says $M = A \otimes {}_AM$.

2.3_* Consider the situation $M = {}_{A'}M_A$ (cf. 1.5_*); $N = {}_AN$. Classically, one turns $X = M \otimes {}_AN$ into a left A'-module by inserting the unique homomorphism $A' \otimes X \to X$ making the following diagram commutative

$$\begin{tikzcd}
A' \otimes M \otimes A \otimes N \arrow[r] \arrow[d, "\tau' \otimes A \otimes N"] & A' \otimes M \otimes N \arrow[r] \arrow[d, "\tau' \otimes N"] & A' \otimes X \arrow[r] & 0 \\
M \otimes A \otimes N \arrow[r] & M \otimes N \arrow[r] & X \arrow[r] & 0
\end{tikzcd}$$

2.3^* In dualising 2.3_*, the failure of $\otimes$ to be left exact imposes a difficulty. However, in the situation $M = {}^{B'}M^B$, $N = {}^BN$, $M \otimes {}^BN$ can be given the structure of a right B'-comodule if *either* B' (i.e., each B'_*) is K-projective, *or* the sequence defining $M \otimes^B N$ is split exact.

2.4$_*$ Consider the situation $(L_{A'}, {}_{A'}M_A, {}_AN)$; then, classically

$$L \otimes_{A'} (M \otimes_A N) \text{ and } (L \otimes_{A'} M) \otimes_A N$$

are naturally isomorphic. If one wants to prove this without mentioning elements, directly from 2.3$_*$, one uses the following lemma, true in any abelian category:

2.41$_*$ **Lemma.** *The diagram of exact rows and columns*

$$\begin{tikzcd}
0 & 0 & & \\
X_1 \arrow[u] \arrow[r] & X_2 \arrow[u] & & \\
Y_1 \arrow[u] \arrow[r] & Y_2 \arrow[u] \arrow[r] & Y_3 \arrow[r] & 0 \\
& Z_2 \arrow[u] \arrow[r] & Z_3 \arrow[u] \arrow[r] & 0
\end{tikzcd}$$

can be completed in a unique way to

$$\begin{tikzcd}
0 & 0 & 0 & \\
X_1 \arrow[u] \arrow[r] & X_2 \arrow[u] \arrow[r] & X_3 \arrow[u] \arrow[r] & 0 \\
Y_1 \arrow[u] \arrow[r] & Y_2 \arrow[u] \arrow[r] & Y_3 \arrow[u] \arrow[r] & 0 \\
& Z_2 \arrow[u] \arrow[r] & Z_3 \arrow[u] \arrow[r] & 0.
\end{tikzcd}$$

2.4* Using the dual lemma in the situation $(L^{B'}, {}^{B'}M^B, {}^BN)$, one proves that $L \otimes^{B'} (M \otimes^B N)$ and $(L \otimes^{B'} M) \otimes^B N$ are naturally isomorphic, provided *either* B', B, L, N are K-projective, *or* the sequences defining $L \otimes^{B'} M$, $M \otimes^B N$ are split exact; we denote either module by $L \otimes^{B'} M \otimes^B N$.

2.5* Consider the situation $(M^B, {}^BN)$, $(M'^{B'}, {}^{B'}N')$, let $f \colon B \to B'$ be a map of G.C. modules and let $m \colon M \to M'$, $n \colon N \to N'$ be f-maps; cf. 1.4*. Then, into the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & M \otimes^B N \arrow[r] & M \otimes N \arrow[r, "\Delta"] \arrow[d, "m \otimes n"] & M \otimes^B \otimes N \arrow[d, "m \otimes f \otimes n"] \\
0 \arrow[r] & M' \otimes^{B'} N' \arrow[r] & M' \otimes N' \arrow[r, "\Delta"] & M' \otimes^{B'} \otimes N'
\end{tikzcd}$$

a unique map

$$m \otimes^* n = m \otimes^f n \colon M \otimes^B N \to M' \otimes^{B'} N'$$

can be inserted; if it exists at all it does not depend on the f used in its construction; the notation $\otimes^f$ is useful as a brief way of indicating what is being used to justify existence.

Under suitable hypotheses, cf. 2.3*, 2.4*, symbols such as $l \otimes^* m \otimes^* n$ are well defined.

2.5$_*$ The dual remarks, leading to symbols such as $m \otimes_* n = m \otimes_f n$, $l \otimes_* m \otimes_* n$, are classical.

3. Some permutations.

Notation. Let A, B be graded modules. By $T\colon A \otimes B \to B \otimes A$ we denote the map of graded modules given by $T(a \otimes b) = (-1)^{pq} b \otimes a$ if $a \in A_p$, $b \in B_q$.

Let $A_1, \cdots, A_k$ be graded modules and $(i_1, \ldots, i_k)$ some permutation of $(1, \cdots, k)$. We shall denote by

$$(i_1, \cdots, i_k) : A_1 \otimes \cdots \otimes A_k \to (A_{i_1} \otimes \cdots \otimes A_{i_k})$$

the map of graded modules that permutes the factors as indicated and introduces the same signs as any succession of T's that produces the same permutation; the uniqueness is clear. Note that $T = (2,1)$.

3.1* *Definition.* Let (B_i, ψ_i) $(i = 1, \cdots, k)$ be co-algebras. We turn $B = B_1 \otimes \cdots \otimes B_k$ into a co-algebra by the definition

$$\psi = (1, 3, \cdots, 2k-1, 2, \cdots, 2k)(\psi_1 \otimes \cdots \otimes \psi_k)$$

Let (N_i, σ_i) be B_i-co-modules. We turn $N = N_1 \otimes \cdots \otimes N_k$ into a B-co-module by the definition

$$\sigma = (1, 3, \cdots, 2k-1, 2, \cdots, 2k)(\sigma_1 \otimes \cdots \otimes \sigma_k)$$

The dual Definition 3.1$_*$ is classical.

3.2$_*$ Let (A, ϕ) be an algebra; $\phi\colon A \otimes A \to A$ is a map of algebras if and only if $\phi = \phi T$ in which case we call A (or ϕ) *commutative.*

The dual statement 3.2* is evident.

3.3$_*$ LEMMA. *Let (A, ϕ) be an algebra, and let $\tau\colon A \otimes M \to M$, $\tau'\colon M \otimes A \to M$ be maps of G.C. modules such that $\tau' = \tau T$. The following conditions are equivalent:*

(i) *(M, τ) is a left A-module and (M, τ') a right A-module.*

(ii) *Either (M, τ) or (M, τ') is an A-module and the condition of* 1.5$_*$ *is satisfied.*

(iii) *(M, τ) is an A-module and $\tau\colon A \otimes M \to M$ is a ϕ-map; cf.* 1.4_*, *where we regard $A \otimes M$ as an $A \otimes A$-module by* 3.1.

(iv) *(M, τ') is an A-module and $\tau'\colon M \otimes A \to M$ is a ϕ-map.*

3.4_* *Definition.* If the conditions of 3.3_* are satisfied we say that τ, τ' are *central*; that M is a *central A-module*, or that the situation ${}_AM_A$ is central.

3.5_* LEMMA. *Let (M, τ) be an A-module. τ is central if A is commutative.*

Remark. The converse is true if

$$A = A \otimes K \xrightarrow{A \otimes \eta} A \otimes M \xrightarrow{\tau} M$$

is a monomorphism.

The duals 3.3^*, 3.4^*, 3.5^* are evident.

3. 6 Consider the situation $(M_A, {}_AN, M'_{A'}, {}_{A'}N')$; it is easily seen, by looking at "elements," that $(M \otimes_A N) \otimes (M' \otimes_{A'} N')$ and

$$(M \otimes N) \otimes_{A \otimes A'} (M' \otimes N')$$

are naturally isomorphic; rewriting the argument in terms of mappings is a bit cumbersome, but there is no essential difficulty; we obtain the following dual lemmas:

3.61_* LEMMA. *Let M_i be a left A_i-module, N_i a right A_i-module; $i = 1, \cdots, k$. There is a unique homomorphism h_*, and it is an isomorphism, such that the following diagram is commutative:*

$$\begin{tikzcd}
M_1 \otimes N_1 \otimes \cdots \otimes M_k \otimes N_k \arrow[r] \arrow[d, "{(1, 3, \cdots, 2k-1, 2, 4, \cdots, 2k)}"] & (M_1 \otimes_{A_1} N_1) \otimes \cdots \otimes (M_k \otimes_{A_k} N_k) \arrow[r] \arrow[d, "h_*"] & 0 \\
{}_1 \otimes \cdots \otimes M_k \otimes N_1 \otimes \cdots \otimes N_k \arrow[r] & (M_1 \otimes \cdots \otimes M_k) \otimes_{A_1 \otimes \cdots \otimes A_k} (N_1 \otimes \cdots \otimes &
\end{tikzcd}$$

3.61^* LEMMA. *Let M_i be a right B_i co-module, N_i a left B_i-co-module, $i = 1, 2, \cdots, k$.*

Suppose either that M_i, N_i, $M_i \otimes^{B_i} N_i$ are all K-projective; or else, that the sequences defining $M_i \otimes^{B_i} N_i$ are all split-exact.

Then there is a unique h^*, and it is an isomorphism, such that the following diagram is commutative

$$\begin{tikzcd}
0 \arrow[r] & (M_1 \otimes^{B_1} N_1) \otimes \cdots \otimes (M_k \otimes^{B_k} N_k) \arrow[r] & M_1 \otimes N_1 \otimes \cdots \otimes M_k \otimes N_k \\
0 \arrow[r] & (M_1 \otimes \cdots \otimes M_k) \otimes^{B_1 \otimes \cdots \otimes B_k} (N_1 \otimes \cdots \otimes N_k) \arrow[r] \arrow[u, "h^*"] & M_1 \otimes \cdots \otimes M_k \otimes N_1 \otimes \cdots \otimes N \arrow[u, "{(1, k+1, \cdots k, 2k)}"']
\end{tikzcd}$$

3.7 The results of 3.6 will now be used to investigate when the tensor product or co-tensor product of two algebras or co-algebras has an induced product or co-product. Thus, letting (A, ϕ) denote an algebra and (B, ψ) a co-algebra, we investigate 4 cases:

I_*: $M = M_A$, $N = {}_A N$, (M, ϕ_M), (N, ϕ_N) are algebras.
II_*: $M = M_A$, $N = {}_A N$, (M, ψ_M), (N, ψ_N) are co-algebras.
I^*: $M = M^B$, $N = {}^B N$, (M, ψ_M), (N, ψ_N) are co-algebras.
II^*: $M = M^B$, $N = {}^B N$, (M, ϕ_M), (N, ϕ_N) are algebras.

We write out explicitly the less familiar cases, I^*, II^*; the dual cases I_*, II_* are very easily handled by "classical" methods.

3.71* We make the following

Assumption. M, N and $M \otimes^B N$ are K-projective; or else, the sequence defining $M \otimes^B N$ is split exact.

I* We use the "co-product"

$$M \otimes^B N \xrightarrow{\psi_M \otimes^* \psi_N} (M \otimes M) \otimes^{B \otimes B} (N \otimes N) \xrightarrow{h_*} (M \otimes^B N) \otimes (M \otimes^B N)$$

which will exist, with $\otimes^* = \otimes^f$ if

CI*: There is a map of G. C. modules $f: B \to B \otimes B$ such that ψ_M, ψ_N are f-maps; (cf. 1.4*).

II* We use the "product"

$$(M \otimes^B N) \otimes (M \otimes^B N) \xrightarrow{h_*^{-1}} (M \otimes M) \otimes^{B \otimes B} (N \otimes N) \xrightarrow{\phi_M \otimes^* \phi_N} (M \otimes^B N)$$

which will exist, with $\otimes^* = \otimes^g$ if

CII*: There is a map of G. C. modules $g: B \otimes B \to B$ such that ϕ_M, ϕ_N are g-maps.

3.72* The investigation of possible choices for f, g in 3.71* seems a bit complicated. We shall restrict ourselves to the following cases:

I*: We take $f = \psi = \psi_B$, and CI* becomes:

DI*: ψ_M, ψ_N are ψ-maps.

II*: We take $g = \phi = \phi_B$ where (B, ϕ) is an algebra, then CII* becomes:

DII*: ϕ_M, ϕ_N are ϕ-maps.

3.8* Condition 3.72*DI*. "ψ_M is a ψ-map" can be expressed more perspicuously. The condition is, of course, the commutativity of

$$\begin{tikzcd}
M\otimes M\otimes B\otimes B & M\otimes B\otimes M\otimes B \arrow[l] & M\otimes M \arrow[l, "\sigma\otimes\sigma"'] \\
M\otimes B \arrow[u, "\psi_M\otimes\psi"'] & & M \arrow[ll, "\sigma"] \arrow[u, "\psi_M"']
\end{tikzcd}$$

3.81* *Definition.* A map of co-algebras $M\xrightarrow{j}B$ is called central if the following diagram is commutative:

$$\begin{tikzcd}
M\otimes B \arrow[rr, leftrightarrow, "T"] & & B\otimes M \\
M\otimes M \arrow[u, "M\otimes j"'] & & M\otimes M \arrow[u, "j\otimes M"'] \\
& M \arrow[ul, "\psi_M"] \arrow[ur, "\psi_M"'] &
\end{tikzcd}$$

3.82* *Notations.* Let $\sigma: M\to B\otimes M$, $\sigma': M\to M\otimes B$ be maps of G. C. modules. We define $j(\sigma)$, $j(\sigma') : M\to B$ as the compositions

$$M\xrightarrow{\sigma} B\otimes M\xrightarrow{B\otimes\epsilon} B\otimes K = B$$

$$M\xrightarrow{\sigma'} M\otimes B\xrightarrow{\epsilon\otimes B} K\otimes B = B.$$

If (M, ψ_M) is a co-algebra and $j: M\to B$ a map of G. C. modules, we define $\sigma(j) : M\to B\otimes M$, $\sigma'(j) : M\to M\otimes B$ as the compositions

$$M\xrightarrow{\psi_M} M\otimes M\xrightarrow{j\otimes M} B\otimes M$$

$$M\xrightarrow{\psi_M} M\otimes M\xrightarrow{M\otimes j} M\otimes B.$$

3.83* LEMMA.

(i) *If* (M,σ) *is a left B-co-module, then* $j(\sigma)$ *is a map of left B-co-modules.*

(ii) *If* $j\colon M \to B$ *is a map of co-algebras, then* $(M,\sigma(j))$ *is a left B-co-module and* $j(\sigma(j)) = j$.

(iii) *If* (M,σ) *is a left B-co-module,* (M,ψ_M) *a co-algebra and* ψ_M *a* ψ*-map, then* $j(\sigma)$ *is a central map of co-algebras,* σ *is central,* (cf. 3.3*), $\sigma = \sigma(j(\sigma))$, $\sigma'(j(\sigma)) = T\sigma$.

(iv) *If* (M,ψ_M) *is a co-algebra and* $j\colon M \to B$ *is a central map of co-algebras, then* $(M,\sigma(j))$ *is a left co-module,* $\sigma(j)$ *is central,* $\sigma'(j) = T\sigma(j)$, *and* ψ_M *is a* ψ*-map.*

From 3.7* we can now read off:

3.84* THEOREM. *Let* $j_M\colon M \to B$, $j_N\colon N \to B$ *be central homomorphisms of co-algebras, so that* $\sigma(j_M)$, $\sigma(j_N)$ *define central situations* ${}^BM^B$, ${}^BN^B$; (cf. 3.4*).

Assume that M, N *and* $M\otimes{}^BN$ *are K-projective; or else that the sequence defining* $M\otimes{}^BN$ *is split exact.*

Then there is a (unique) homomorphism ψ^* *making the following diagram commutative:*

$$\begin{tikzcd}
M\otimes N\otimes M\otimes N & M\otimes N \arrow[l, "\psi_{M\otimes N}"'] \\
(M\otimes{}^BN)\otimes(M\otimes{}^BN) \arrow[u] & (M\otimes{}^BN) \arrow[l, "\psi^*"'] \arrow[u] \\
0 \arrow[u] & 0 \arrow[u]
\end{tikzcd}$$

$(M\otimes{}^BN,\psi^*)$ *is a co-algebra.*

3.8$_*$ The dual of 3.81*, the notion of a central map of algebras, is classical. Corresponding to $\tau\colon A\otimes M \to M$ we define $i(\tau)$ as the composition

$$A = A\otimes K \xrightarrow{A\otimes\eta} A\otimes M \xrightarrow{\tau} M$$

and corresponding to $i\colon A \to M$, where (M,ϕ_M) is an algebra, we define $\tau(i)$ as the composition

$$A\otimes M \xrightarrow{i\otimes M} M\otimes M \xrightarrow{\phi_M} M.$$

$i(\tau)$, $\tau(i)$ are dual to $j(\sigma)$, $\sigma(j)$ of 3.82*.

3.84$_*$ THEOREM. *Let* $i_M : A \to M$, $i_N : A \to N$ *be central maps of algebras, so that* $\tau(i_M)$, $\tau(i_N)$ *define the central situations* ${}_AM_A$, ${}_AN_A$; (cf. 3.4$_*$).

Then there is a (unique) homomorphism ϕ_* *making the following diagram commutative:*

$$\begin{tikzcd}
M \otimes N \otimes M \otimes N \arrow[r, "\phi_{M \otimes N}"'] \arrow[d] & M \otimes N \arrow[d] \\
(M \otimes_A N) \otimes (M \otimes_A N) \arrow[r, "\phi_*"'] \arrow[d] & M \otimes_A N \arrow[d] \\
0 & 0
\end{tikzcd}$$

$(M \otimes_A N, \phi_*)$ *is an algebra.*

4. Hopf algebras. The conditions DII$_*$, DII* of Section 3.7 cannot be re-interpreted in the manner of 3.8 if we remain in the context of 3.7; the main purpose of this chapter is to show that such a re-interpretation becomes possible in a self-dual situation.

4.1 *Definition.* The triple (H, ϕ, ψ) is called a Hopf-algebra if

(i$_*$) (H, ϕ) is an algebra.

(i*) (H, ψ) is a co-algebra.

(ii) The following diagram is commutative:

$$\begin{tikzcd}
H \otimes H \arrow[r, "\phi"] \arrow[d, "\psi \otimes \psi"] & H \arrow[r, "\psi"] & H \otimes H \\
H \otimes H \otimes H \otimes H \arrow[rr, "H \otimes T \otimes H"] & & H \otimes H \otimes H \otimes H \arrow[u, "\phi \otimes \phi"']
\end{tikzcd}$$

Condition (ii) is self-dual and can be interpreted either as "ψ is a homomorphism of algebras" or as "ϕ is a homomorphism of co-algebras." We refer to it as the Hopf-condition; cf. [1].

A homomorphism $f : (H, \phi, \psi) \to (H', \phi', \psi')$ of Hopf algebras is one which is a homomorphism of algebras and of co-algebras.

If H, H' are Hopf-algebras, we define the Hopf-algebra $H \otimes H'$ by combining 3.1$_*$ and 3.1*; the Hopf-condition is easily verified.

4.2$_*$ Let us now be in the situation $(M_A, {}_AN)$, where (A, ϕ) is an algebra having a co-product ψ (but we do not, for the moment, assume the Hopf-condition), and (M, ϕ_M, ψ_M), (N, ϕ_N, ψ_N) are Hopf-algebras. We want to turn $M \otimes_A N$ into a Hopf-algebra; sufficient conditions are easily read off from 3.7 namely DI_*, DII_*. We reformulate DI_* as in 3.8_* by requiring that

$$i = i(\tau) : A \to M \text{ or } N$$

both be central homomorphisms of algebras.

Now we express DII_* one of which is "ψ_M is a ψ-map"; this means the commutativity of

$$\begin{tikzcd}
M \otimes A \arrow[rr, "\tau"] & & M \\
M \otimes M \otimes A \otimes A \arrow[u, "\psi_M \otimes \psi"'] \arrow[r, "M \otimes \tau \otimes A"] & M \otimes A \otimes M \otimes A \arrow[r, "\tau \otimes \tau"] & M \otimes M \arrow[u, "\psi_M"']
\end{tikzcd}$$

If we substitute $\tau = \tau(i) = \phi_M(M \otimes i)$ and use the Hopf-condition for M it is easily seen that the commutativity of the above diagram is equivalent to the requirement that $i \colon A \to M$ is a map of co-algebras.

We summarize:

4.21_* THEOREM. *Let* (M, ϕ_M, ψ_M), (N, ϕ_N, ψ_N) *be Hopf algebras, let* (A, ϕ) *be an algebra with a co-multiplication* ψ.

Let $i_M \colon A \to M$, $i_N \colon A \to N$ *be central homomorphisms of algebras and homomorphisms of co-algebras.* $\tau(i_N)$, $\tau(i_M)$ *define the central situations* ${}_AM_A$, ${}_AN_A$ (cf. 3.8_*, 3.4_*).

Then $(M \otimes_A N, \phi_*, \psi_*)$ *is a Hopf-algebra, where* ϕ_*, ψ_* *make commutative diagrams analogous to that in* 3.94_*.

Remark. If either i_M or i_N is a monomorphism it follows that A is a Hopf-algebra with commutative ϕ, and i_M, i_N are homomorphisms of Hopf-algebras.

Proof. In view of what precedes, the only thing that remains to be proved is the Hopf-condition for $M \otimes_A N$. Now, the Hopf-condition for $M \otimes N$ asserts the equality of two homomorphisms $M \otimes N \otimes M \otimes N \to M \otimes N \otimes M \otimes N$; the equality of the induced homomorphisms

$$(M \otimes_A N) \otimes (M \otimes_A N) \to (M \otimes_A N) \otimes (M \otimes_A N)$$

follows from the uniqueness of the homomorphisms "inserted" in 2.6_*, 3.6_*; the Hopf-condition for $M \otimes N$ "maps onto" that for $M \otimes_A N$.

4.21* THEOREM. *Let* (M,ϕ_M,ψ_M), (N,ϕ_N,ψ_N) *be Hopf-algebras, let* (B,ψ) *be a co-algebra with multiplication* ϕ.

Let $j_M\colon M\to B$, $j_N\colon N\to B$ *be central homomorphisms of co-algebras and homomorphisms of algebras.* $\sigma(j_N)$, $\sigma(j_M)$ *define the central situations* ${}^BM^B$, ${}^BN^B$.

Suppose, moreover, that M, N *and* $M\otimes{}^BN$ *are K-projective; or else that the sequence defining* $M\otimes{}^BN$ *is split exact. Then* $(M\otimes{}^BN,\phi^*,\psi^*)$ *is Hopf-algebra, where* ϕ^*, ψ^* *make commutative diagrams analogous to that in* 3.94*.

Remark. If one of j_M, j_N is onto, it follows that B is a Hopf-algebra with commutative ψ and j_M, j_N are homomorphisms of Hopf-algebras.

5. Extensions of algebras and co-algebras. The grading of our various modules has played no part so far; now it will. If $J=\{J_n\}$ is a graded module, we shall often, and ambiguously use J_n to denote the graded module that is zero in dimensions other than n, and J_n in that dimension. By $q_n=q_n(J)\colon J_n\to J$, $p_n=p_n(J)\colon J\to J_n$ we denote the injection and projection.

5.0 LEMMA. *Let* J, L, M, N *be G.C. modules, and* $f\colon J\otimes L\to M\otimes N$ *a map of G.C. modules such that* $(M\otimes p_n)f(J\otimes q_n)\colon J\otimes L_n\to M\otimes N_n$ *is an isomorphism for* $n\geqq 0$. *Then* f *is an isomorphism.*

This lemma is well known from the theory of spectral sequences.

5.1 Let A, B be G.C. modules. We define $i=i_A\colon A\to A\otimes B$ as the map $A=A\otimes K\xrightarrow{A\otimes\eta}A\otimes B$, and $j=j_B\colon A\otimes B\to B$ as the map $A\otimes B\xrightarrow{\epsilon\otimes B}K\otimes B=B$. We refer to

$$A\xrightarrow{i_A}A\otimes B\xrightarrow{j_B}B$$

as the basic (A,B)-sequence, and denote it by $E(A,B)$. The map $f\colon A\otimes B\to A\otimes B$ will be called *admissible* if

$$\begin{tikzcd}
 & A\otimes B \arrow[dd, "f"] \arrow[dr, "j"] & \\
A \arrow[ur, "i"] \arrow[dr, "i"'] & & B \arrow[dl, "j"] \\
 & A\otimes B &
\end{tikzcd}$$

is commutative.

5.11_* LEMMA. *Let* A *be an algebra and give to* $A\otimes B$ *the evident* *structure of a left A-module. If f is admissible and a map of A-modules, then it is an isomorphism.*

5.11^* LEMMA. *Let B be a co-algebra and give $A\otimes B$ the evident structure of a right B-co-module; if f is an admissible map of B-co-modules, then it is an isomorphism.*

We prove 5.11_* by proving that $(A\otimes p_n)f(A\otimes q_n): A\otimes B_n \to A\otimes B_n$ is the identity and then applying 5.0.

Indeed, let $b\in B_n$. Then, since f is admissible

$$f(1\otimes b) = 1\otimes b + \textstyle\sum_\alpha a_\alpha \otimes b_\alpha$$

where $\dim a_\alpha > 0$, $\dim a_\alpha + \dim b_\alpha = n$. Hence $(A\otimes p_n)f(1\otimes b) = 1\otimes b$. Hence $(A\otimes p_n)f(a\otimes b) = a\otimes b$ because f is a map of A-modules.

To prove 5.11^* one first puts this argument into an "element-free form"—after which dualization is mechanical. The details are surprisingly awkward!

5.2 *Notation.* A commutative diagram of G. C. modules

$$\begin{tikzcd}
A \arrow[r, "i"] \arrow[d, "a"] & C \arrow[r, "j"] \arrow[d, "c"] & B \arrow[d, "b"] \\
A' \arrow[r, "i'"] & C' \arrow[r, "j'"] & B'
\end{tikzcd}$$

will be referred to as the map (a, c, b); a complete notation would be $(a, c, b): (i, j) \to (i', j')$. Composition of such maps is defined in the obvious way.

5.3_* *Definition.* A sequence of homomorphisms of algebras

$$E_*:\ A \xrightarrow{i} C \xrightarrow{j} B$$

will be called an (A,B)-*extension* of algebras if there is an isomorphism of left A-modules $\Gamma: (C, \tau(i)) \to (A\otimes B, \tau(i_A))$, cf. 3.83_*, 5.1, such that $(A, \Gamma, B): E_* \to E(A,B)$ is a map.

Observe that $ji = \eta(B)\epsilon(A)$.

A map (a, c, b) in which a, c, b are maps of algebras will be called a map of extensions of algebras.

A map of extensions (A, c, B) will be called an *equivalence* (of (A,B)-extensions). In this case, c is an isomorphism, by 5.11_*; showing that "equivalence" does indeed define an equivalence relation.

If i is a central map of algebras, 3.81_*, we say that E_* is a *central* extension of algebras. In this case A is commutative.

5.3* The dual definition for an *(A, B)-extension of co-algebras*, $E^*: A \xrightarrow{i} C \xrightarrow{j} B$ is evident. In this case we require that Γ is an isomorphism of right B-comodules. E^* will be called central if j is.

5.4$_*$ THEOREM. *Let $E_*: A \xrightarrow{i} C \xrightarrow{j} B$ be a central extension of algebras and $a: A \rightarrow A'$ a central map of algebras.*

There exists an (A', B)-extension of algebras

$$a_*E_*: A' \rightarrow C_* \rightarrow B$$

and a map of extensions of algebras $(a, a_*, B): E_* \rightarrow a_*E_*$ satisfying the following condition:

Given a map of extensions $(a, c, b): E_* \rightarrow E'_*$, there exists a map $(A', c_*, b): a_*E_* \rightarrow E'_*$ such that

$$(a, c, b) = (A', c_*, b)\,(a, a_*, B)$$

The extension a_*E_* is unique up to an equivalence; a_*E_* having been chosen, a_* is unique.

Proof. (For an alternative proof, see 8.3$_*$)

The uniqueness follows by a familiar argument; it remains to establish existence.

The central map i defines the central situation ${}_AC_A$, cf. 3.83$_*$; similarly a defines ${}_AA'_A$. We define $C_* = A' \otimes_A C$; it is an algebra by 3.84$_*$; and

$$\begin{tikzcd} A' = A' \otimes_A A \arrow[r, "A' \otimes_* i"'] & A' \otimes_A C \arrow[r, "\epsilon \otimes_* j"'] & K \otimes B = B \end{tikzcd}$$

defines a_*E_*. The map $A' \otimes_* i$ is defined because i is an A-map; the symbol $\epsilon \otimes_* j$ requires some explanation! We regard K, B as A-modules with "trivial" operation, i. e.

$K \otimes A \rightarrow A$ is defined by $k \otimes a \rightarrow k\epsilon(a)$
$A \otimes B \rightarrow B$ is defined by $a \otimes b \rightarrow \epsilon(a)b$.

It is then clear that $K \otimes B = K \otimes_A B$, and that ϵ, j are A-maps. It is easily seen that $A' \otimes_* i$, $\epsilon \otimes_* j$ are maps of algebras, cf. 3.8$_*$

$$a_*: C \rightarrow C_* \text{ is defined as}$$

$a \otimes_* C: A \otimes_A C \rightarrow A' \otimes_A C$; this is possible since a, C are A-maps, cf. 1.4$_*$.

To see that what we have defined is, indeed, an (A', B)-extension, consider first the diagram

$$\begin{tikzcd}
A'\otimes A\otimes C \arrow[r, "A'\otimes A\otimes\Gamma"] \arrow[d, "\Delta"] & A'\otimes A\otimes A\otimes B \arrow[d, "\Delta\otimes B"] \\
A'\otimes C \arrow[r, "A'\otimes\Gamma"] \arrow[d] & A'\otimes A\otimes B \arrow[d] \\
A'\otimes_A C \arrow[r, "\Gamma'"] \arrow[d] & A'\otimes B \arrow[d] \\
0 & 0
\end{tikzcd}$$

into which Γ' can be inserted uniquely; it is clearly a map of A'-modules; and since Γ is an isomorphism, so is Γ'.

Next, we consider the diagram

$$\begin{tikzcd}
A'\otimes A \arrow[rrrr, "="] \arrow[dr] \arrow[dddd, swap, "\tau"] & & & & A'\otimes A\otimes K \arrow[dl] \arrow[dddd, "\tau\otimes K"] \\
& A'\otimes C \arrow[rr] \arrow[dr] \arrow[dd] & & A'\otimes A\otimes B \arrow[dl] \arrow[dd] & \\
& & K\otimes B = K\otimes K\otimes B & & \\
& A'\otimes_A C \arrow[ur] \arrow[rr, swap, "\Gamma'"] & & A'\otimes_A A\otimes B \arrow[ul] & \\
A' = A'\otimes_A A \arrow[ur] \arrow[rrrr, leftrightarrow] & & & & A'\otimes K \arrow[ul]
\end{tikzcd}$$

from which it follows that (A', Γ', B) is a map.

Now, let $(a, c, b) : E_* \to E'_*$, where E'_* is $A' \to C' \to B'$, be given. We define $c_* : C_* \to C'$ by

$$c_* = A'\otimes_* c : A'\otimes_A C \to A'\otimes_{A'} C' = C';$$

the map $A'\otimes_* c$ is defined as $A'\otimes_a c$, cf. 2. 6_*; this is possible since A', c are a-maps, cf. 1. 4_*: The verifications are trivial. Equally trivial are the verifications that (A', c_*, b) is a map, and that $c_* a_* = (A'\otimes_* c)\,(a\otimes_* C) = a\otimes_* c : A\otimes_A C \to A'\otimes_{A'} C'$ *is* c.

This completes the proof.

5.4* THEOREM. *Let* $E^* : A \xrightarrow{i} C \xrightarrow{j} B$ *be a central extension of co-algebras and let* $b : B' \to B$ *be a central map of co-algebras.*

There exists an (A, B')*-extension of co-algebras*

$$b^*E : A \to C^* \to B'$$

and a map of extensions

$$(A, b^*, b) : b^*E^* \to E$$

(so that b^* is a map $C^* \to C$) satisfying the following condition:

Given a map of extensions $(a, c, b) : E^{*\prime} \to E^*$, there exists a map $(a, c^*, B') : E^{*\prime} \to b^*E^*$ such that $(a, c, b) = (A, b^*, b)(a, c^*, B')$.

The extension b^*E^* having been chosen, b^* is unique.

Remark. At first sight it would seem that, as usual, we have to add an assumption; the fact that this is not so is due to the *split* exactness of 2.2^*; we thus obtain the dual of the first diagram in the proof of 5.4_*.

5.5_* Given an algebra A we can regard K as an A-module by the definition $\tau : K \otimes A \to K$, $\tau = K \otimes \epsilon$. For any A-module M, $K \otimes_A M$ is defined by

$$A \otimes M \xrightarrow{\Delta} M \to K \otimes_A M \to 0$$

where $\Delta = \epsilon \otimes M - \tau$. From 2.2_* we have the split exact sequence

$$A \otimes A \xrightarrow{\Delta} A \xrightarrow{\epsilon} K \to 0$$

whence

$$A \otimes (A \otimes B) \xrightarrow{\Delta} A \otimes B \xrightarrow{j_B} B \to 0$$

is split exact. Hence

5.51_* LEMMA. *If $A \xrightarrow{i} C \xrightarrow{j} B$ is an extension of algebras, then*

$$A \otimes C \xrightarrow{\Delta} C \xrightarrow{j} B \to 0$$

is split exact; in particular $B = K \otimes_A C$.

The following lemma provides a partial converse:

5.52_* LEMMA. *Let i_1, j_1, i_2, j_2 be maps of algebras such that*

$$\begin{tikzcd} & C_1 \arrow[dd, "f"] \arrow[dr, "j_1"] & \\ A \arrow[ur, "i_1"] \arrow[dr, "i_2"'] & & B \\ & C_2 \arrow[ur, "j_2"'] & \end{tikzcd}$$

is a commutative diagram, where f is a map of left A-modules; we regard C_t $(t = 1, 2)$ as left A-modules, using $\tau(i)$. Suppose that the sequences

$$A \otimes C_t \xrightarrow{\Delta} C_t \xrightarrow{j} B \to 0$$

are exact.

Then f is an epimorphism.

Proof. Let V be the co-kernel of f. We obtain the following diagram of exact rows and columns

$$\begin{tikzcd}
A \otimes C_1 \arrow[r] \arrow[d, "A \otimes f"] & C_1 \arrow[r] \arrow[d, "f"] & B \arrow[r] \arrow[d, equal] & 0 \\
A \otimes C_2 \arrow[r] \arrow[d] & C_2 \arrow[r] \arrow[d] & B \arrow[r] & 0 \\
A \otimes V \arrow[r] \arrow[d] & V \arrow[d] & & \\
0 & 0 & &
\end{tikzcd}$$

It follows from 2.4$_*$ that $A \otimes V \rightarrow V$ is onto. Now, introducing dimensions, $A_0 \otimes V$ maps into zero, hence $\sum A_k \otimes V_{n-k}$ for $k > 0$ maps onto V_n; but $V_0 = 0$; hence, by induction, $V_n = 0$.

5.5* Dually, we can regard K as a B-comodule with $\sigma\colon K \rightarrow B \otimes K$ defined by $\sigma = \eta \otimes K$; for any $M = M^B$, $M \otimes {}^B K$ is given by

$$0 \rightarrow M \otimes {}^B K \rightarrow M \xrightarrow{\Delta} B \otimes M$$

where $\Delta = \eta \otimes M - \sigma$. Using 2.2* we prove

5.51* LEMMA. *If* $A \xrightarrow{i} C \xrightarrow{j} B$ *is an extension of co-algebras, then*

$$0 \rightarrow A \xrightarrow{i} C \xrightarrow{\Delta} C \otimes B$$

is split exact. In particular, $A = C \otimes {}^B K$.

5.52* LEMMA. *In the notation of* 5.52$_*$, *let* i_1, i_2, j_1, j_2 *be maps of co-algebras and* f *a map of right* B*-comodules. Suppose that*

Either

$0 \rightarrow U \rightarrow C_1 \xrightarrow{f} C_2$ *is split exact*

Or

B *is* K*-projective*

and that

$0 \rightarrow A \xrightarrow{i} C_t \xrightarrow{\Delta} C_t \otimes B$ $(t = 1, 2)$ *are exact.*

Then f *is a monomorphism.*

5.6 If it were true that by supposing the sequences given in 5.52$_*$, 5.52* to be split exact, we could prove f in each case to be an *isomorphism,* a more convenient definition of "extension"—avoiding the non-invariant Γ—could be used. Using Lemma 7.1, one can *almost* do this; but some necessary extra assumptions *on the mappings* seem to make this approach impossible.

6. Extensions of Hopf-algebras.

6.1 *Definition.* A sequence of homomorphisms of Hopf-algebras

$$E: A \xrightarrow{i} C \xrightarrow{j} B$$

is called an (A, B)-extension of Hopf-algebras if there is an isomorphism of left A-modules and of right B-comodules

$$\Gamma: (C, \tau(i), \sigma(j)) \to (A \otimes B, \tau(i_A), \sigma(j_B))$$

such that $(A, \Gamma, B): E \to E(A, B)$ is a map; cf. 3.93_*, 3.93^*, 5.1.

A map $(a, c, b): E \to E'$ in which a, c, b are maps of Hopf-algebras will be called a map of extensions of Hopf-algebras; a map of extensions (A, c, B) will be called an equivalence; in this case c is an isomorphism; the last statement is true for 2 reasons; this sort of logical redundancy will occur several times in the theory of extensions of Hopf-algebras.

If i is a central map of algebras and j a central map of co-algebras, we call E a central extension of Hopf-algebras.

6.2 Since an extension of Hopf-algebras is both an extension of algebras and one of co-algebras, all the last chapter is applicable; in particular with $E: A \to C \to B$ we can associate two split exact sequences, namely

$$A \otimes C \xrightarrow{\Delta} C \xrightarrow{j} B \to 0 \qquad \text{———} (1_*)$$

$$0 \to A \xrightarrow{i} C \xrightarrow{\Delta} C \otimes B \qquad \text{———} (1^*)$$

cf. 5.51_*, 5.51^*.

6.21 LEMMA. *In the notation of 5.52_*, let i_1, i_2, j_1, j_2 be maps of Hopf-algebras, and f a map of left A-modules and right B-comodules. Suppose that*

either $0 \to U \to C_1 \xrightarrow{f} C_2$ *is split exact*
or B *is K-projective*

and that the sequences 1_*, 1^* above are exact with $C = C_t$, $t = 1, 2$.

Then f is an isomorphism.

Proof. 5.52_*, 5.52^*.

6.3_* THEOREM. *Let* $E: A \xrightarrow{i} C \xrightarrow{j} B$ *be an extension of Hopf-algebras, and let* i *be a central map of algebras; also let* $a: A \rightarrow A'$ *be a map of Hopf-algebras which is a central map of algebras.*

There exists an (A', B)*-extension of Hopf-algebras* a_*E *satisfying the condition of* 5.4_*, *where all maps are now maps of Hopf-algebras.*

Proof. The facts additional to those of 5.4_* follow easily from 4.21_*. For an alternative proof, see 8.7.

6.3^* THEOREM. *Let* $E: A \xrightarrow{i} C \xrightarrow{j} B$ *be an extension of Hopf-algebras and let* j *be a central map of co-algebras; also let* $b: B' \rightarrow B$ *be a map of Hopf-algebras which is a central map of co-algebras.*

There exists an (A, B') *extension of Hopf-algebras* b^*E *satisfying the conditions of* 5.4^*, *where all maps are now maps of Hopf-algebras.*

6.4 THEOREM. *Let* $E: A \xrightarrow{i} C \xrightarrow{j} B$ *be a central extension of Hopf-algebras. Let the maps of Hopf-algebras* $a: A \rightarrow A'$, $b: B' \rightarrow B$ *be central maps of algebras, co-algebras respectively.*

Then $a_*(b^*E)$ *and* $b^*(a_*E)$ *are equivalent extensions.*

Proof. Consider the diagram

$$\begin{tikzcd}
b^*E \arrow[r, "(A, \cdot, b)"] \arrow[d, "(a, \cdot, B')"] & E \arrow[r, "(a, \cdot, B)"] & a_*E \\
a_*(b^*E) & & b^*(a_*E) \arrow[u, "(A', \cdot, b)"']
\end{tikzcd}$$

where $\cdot$ denotes various maps of Hopf-algebras that we do not name.

By the defining property of $b^*(a_*E)$ we can insert a map of extensions $(a, \cdot, B'): b^*E \rightarrow b^*(a_*E)$ which makes the diagram commutative.

By the defininig property of $a_*(b^*E)$ we can now insert a map $(A', \cdot B')$: $a_*(b^*E) \rightarrow b^*(a_*E)$ which makes the diagram commutative.

This map is the asserted equivalence.

Remark. Dually, we can first insert a map $(A, \cdot, b): a_*b^*E \rightarrow a_*E$, and then a map $a_*b^*E \rightarrow b^*a_*E$.

For another proof, see 8.7.

By $a_*b^*E = b^*a_*E$ we shall from now on denote *some* extension in the equivalence class defined by Theorem 6.4.

6.5 *Definition.* Let

$$E: A \xrightarrow{i} C \xrightarrow{j} B$$
$$E': A' \xrightarrow{i'} C' \xrightarrow{j'} B'$$

be extensions of algebras, co-algebras or Hopf-algebras. It is easily seen that so is

$$A \otimes A' \xrightarrow{i \otimes i'} C \otimes C' \xrightarrow{j \otimes j'} B \otimes B'.$$

We denote it, in each case, by $E \otimes E'$.

The extension class of E, E' uniquely determines that of $E \otimes E'$.

6.6 *Definition.* Let A, B be Hopf-algebras and let A be a commutative algebra, B a commutative co-algebra. By $\mathcal{E}(A,B)$ we shall denote the set of equivalence classes of central (A,B)-extensions of Hopf-algebras.

Observe that $\phi: A \otimes A \to A$, $\psi: B \to B \otimes B$ are maps of Hopf-algebras.

This allows us to introduce in $\mathcal{E}(A,B)$ a binary operation $+$ as follows: Let $x, y \in \mathcal{E}(A,B)$; let E_t be an extension in the class $t = x, y$. We define $x + y$ as the class of

$$\phi_* \psi^* (E_x \otimes E_y) = \psi^* \phi_* (E_x \otimes E_y)$$

6.61 THEOREM. *$\mathcal{E}(A,B)$ with the operation $+$ is a commutative group. The extension class of* $E(A,B): A \xrightarrow{i_A} A \otimes B \xrightarrow{j_B} B$ (cf. 5.1), *where $A \otimes B$ is given the tensor product Hopf-algebra structure, is the zero element of this group.*

This theorem will be proved in Section 8. The only real difficulty is to prove the existence of inverses.

Remark. Analogy with the classical theory of Baer suggests that the inverse of $A \xrightarrow{i} C \xrightarrow{j} B$ might be

$$A \xrightarrow{ic} C \xrightarrow{j} B \quad \text{or} \quad A \xrightarrow{i} C \xrightarrow{cj} B,$$

where $c: A \to A$ or $B \to B$ is the "conjugation map," cf. [1]. Examples show that this is *not* so.

7. **Bitensor products.** We shall give brief indications of an alternative

approach to theorems 6.4, 6.6. The functor we introduce seems to be of some independent interest, but will not be pursued here.

We begin with a self-dual analogue to 2.4_*, 2.4^*.

7.1 LEMMA. *The diagram of exact rows and columns*

$$\begin{tikzcd}
& & 0 \arrow[d] \\
& X_1 \arrow[r] \arrow[d] & X_2 \arrow[d] \\
& Y_1 \arrow[r] \arrow[d] & Y_2 \arrow[r] \arrow[d] & Y_3 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] & Z_3 \\
& 0
\end{tikzcd}$$

can be completed in a unique way to

$$\begin{tikzcd}
& & 0 \arrow[d] & 0 \arrow[d] \\
& X_1 \arrow[r] \arrow[d] & X_2 \arrow[r] \arrow[d] & X_3 \arrow[r] \arrow[d] & 0 \\
& Y_1 \arrow[r] \arrow[d] & Y_2 \arrow[r] \arrow[d] & Y_3 \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] & Z_3 \\
& 0
\end{tikzcd}$$

Now, let A be an algebra, B a co-algebra and let $L = L_A = L^B$ and let $M = {}_AM = {}^BM$, in other words L, M are both A-modules and B-comodules. Consider the diagram

$$\begin{tikzcd}
& 0 \arrow[d] & 0 \arrow[d] \\
& U \arrow[r] \arrow[d] & L \otimes {}^BM & * \\
& L \otimes A \otimes M \arrow[r, "\Delta"] \arrow[d] & L \otimes M \arrow[r] \arrow[d, "\Delta'"] \arrow[from=u] & L \otimes {}_AM \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & I \arrow[r] \arrow[d] & L \otimes B \otimes M \arrow[r] & V \arrow[r] & 0 \\
& 0
\end{tikzcd}$$

in which I is defined as image $\Delta'\Delta$, and U, V are the kernel and co-kernel as indicated. The unique module which can be inserted at * is denoted by $L \otimes {}^B{}_AM$ and called the bi-tensor product of L and M. Observe that if either the operations or co-operations are "trivial," then $L \otimes {}^B{}_AM$ is $L \otimes {}^BM$ or $L \otimes {}_AM$. Thus, the theory of the bitensor product subsumes both our dual theories; unfortunately the diagrams one is led to in order to prove statements analogous to Chapters 2 and 3 above are very big and hard to manage.

Now, let us be in the situation of Theorem 6.4; it will be apparent in the next chapter that both $a_*(b^*E)$ and $b^*(a_*E)$ can be taken as

$$A' \to C \otimes^{B}{}_{A}(A' \otimes B') \to B',$$

the isomorphism of 6.4 being the one implicit in Lemma 7.1.

In 6.6, if E_t is the extension

$$A \to C_t \to B \qquad (t = x, y)$$

then $x + y$ is the class of

$$A \to C_x \otimes^{B}{}_{A} C_y \to B.$$

8. Twisted tensor products.

The main purpose of the present chapter is the proof of Theorem 6.61; for this we develop a second method of dealing with central extensions; the idea is simple! We exploit the isomorphisms Γ appearing in definitions 5.3_*, 5.3^*, 6.1 and thus choose for each extension the representative $E(A, B)$, 5.1. Each extension will be, essentially, the module $A \otimes B$—but with a multiplication or co-multiplication different—"twisted"—from the standard one; we describe certain functions which define the twist; the resulting non-invariant formalism is far more explicit than the preceding theory.

Many formulas will be asserted without proof; the proof in each case is a commutative diagram—and usually a big one.

8.0 Throughout the chapter we keep fixed the notations of 5.1, as well as

$$j_A\colon A \otimes B \to A \text{ defined by } j_A = A \otimes \epsilon$$
$$i_B\colon B \to A \otimes B \text{ defined by } i_B = \eta \otimes B$$

where we use the usual identifications $A = A \otimes K$, $B = K \otimes B$. Also, we shall write $C = A \otimes B$, so that, as G. C. modules, all our extensions will be taken in the form $A \xrightarrow{i_A} C \xrightarrow{j_B} B$; the maps i_A, j_B belong naturally to the extensions; the maps i_B, j_A are part of the chosen "coordinate system" and are, in general, *not* compatible with the various structures.

8.1_* *Definition.* $(A \otimes B, \phi, \phi_A, \phi_B) = (C, \phi, \phi_A, \phi_B)$ or briefly (C, ϕ) will be called a twisted tensor-product of algebras (T. T. A.) if

(i) (C, ϕ), (A, ϕ_A), (B, ϕ_B) are algebras

(ii) The following diagrams are commutative

$$\begin{tikzcd}
A\otimes A\otimes B \arrow[r, "i_A\otimes C"] \arrow[dr, "\phi_A\otimes B"'] & C\otimes C \arrow[d, "\phi"] \\
& C
\end{tikzcd}$$

(1)

$$\begin{tikzcd}
A\otimes B\otimes A \arrow[r, "C\otimes i_A"] \arrow[d, "A\otimes T"] & C\otimes C \arrow[d, "\phi"] \\
A\otimes A\otimes B \arrow[r, "\phi_A\otimes B"'] & C
\end{tikzcd}$$

(2)

$$\begin{tikzcd}
C\otimes C \arrow[r, "j_B\otimes j_B"] \arrow[d, "\phi"] & B\otimes B \arrow[d, "\phi_B"] \\
C \arrow[r, "j_B"'] & B
\end{tikzcd}$$

(3)

Remarks. (1) expresses the fact that $\phi_A\otimes B$ is the operation $\tau(i_A)$, cf. 3.92_*; (2) expresses the fact that $(\phi_A\otimes B)(A\otimes T)$ is $\tau'(i_A)$; (3) expresses that j_B is a map of algebras.

It follows easily that i_A is a map of algebras.

8.10_* In proving formulas about T. T. A. it is usually possible to use the "classical" method of using elements. For this purpose the following notation has been found convenient: If c, $c'\in C$ we denote by cc' the "untwisted" product in $A\otimes B$ regarded as an algebra in the canonical manner; and we write $c\circ c' = \phi(c\otimes c')$; furthermore we identify $a\in A$ with $i_Aa = a\otimes 1$ and $b\in B$ with $i_Bb = 1\otimes b$. Now conditions (1), (2), (3) of 8.1_* can be written as follows:

$$\left.\begin{array}{ll}(1) & a\circ c = ac\\ (2) & c\circ a = ca\end{array}\right\}\text{ for } a\in A,\ c\in C$$

$$(3)\quad b\circ b' = bb' + \textstyle\sum_i a^i b^i + a(b,b') \text{ where } \begin{array}{l} a^i, a(b,b')\in A - A_0\\ b^i\in B - B_0.\end{array}$$

If one tries to express the associativity of ϕ using the notation of (3) above, it proves somewhat intractable.

There is no section 8.10^*.

8.11_* LEMMA. *If ϕ_A is commutative and (C,ϕ) is a T. T. A., then $E(A,B)$ is a central extension of algebras.*

Conversely, given a central extension of algebras, cf. 5.3_*, if we use Γ to "transfer" the multiplication of C to $A\otimes B$, we obtain a T. T. A.

Thus, the theory of T. T. A. includes that of central extensions; in a certain sense it is more general! We do not need to assume that ϕ_A is commutative; although, as will be seen, this is still needed in the main theorems. The situation with A not commutative is hard to specify invariantly; it means that C as an A-module, has a set of generators—namely $i_B B$—which anticommute with the elements of $i_A A$.

8.1* *Definition.* $(A\otimes B, \psi, \psi_A, \psi_B) = (C, \psi, \psi_A, \psi_B)$ or briefly (C, ψ) will be called a twisted tensor product of co-algebras (T. T. C.) if

(i) (C, ψ), (A, ψ_A), (B, ψ_B) are co-algebras

(ii) The following diagrams are commutative:

$$\begin{tikzcd}
A\otimes B\otimes B & C\otimes C \arrow[l, "C\otimes j_B"'] \\
 & C \arrow[u, "\psi"'] \arrow[ul, "A\otimes\psi_B"]
\end{tikzcd}$$

(1)

$$\begin{tikzcd}
B\otimes A\otimes B & C\otimes C \arrow[l, "j_B\otimes C"'] \arrow[d, "\psi"] \\
A\otimes B\otimes B \arrow[u, "T\otimes B"'] & C \arrow[l, "A\otimes\psi_B"]
\end{tikzcd}$$

(2)

$$\begin{tikzcd}
C\otimes C & A\otimes A \arrow[l, "i_A\otimes i_A"'] \\
C \arrow[u, "\psi"'] & A \arrow[l, "i_A"] \arrow[u, "\psi_A"']
\end{tikzcd}$$

(3)

It follows easily that j_B also is a map of co-algebras.

8.11* Lemma. *If ψ_B is commutative and (C, ψ) is a T. T. C., then $E(A, B)$ is a central extension of co-algebras.*

Conversely, given a central extension of co-algebras, cf. 5.3, if we use Γ to "transfer" the co-multiplication of C to $A\otimes B$, we obtain a T. T. C.*

Convention. In order to abbreviate enunciations, (A, ϕ_A), (B, ϕ_B) will always denote algebras, (A, ψ_A), (B, ψ_B) co-algebras.

8.2$_*$ Lemma. *If (C, ϕ) is a T. T. A. and $\lambda = \phi(i_B\otimes i_B) : B\otimes B \to C$ then*

$$\phi = (\phi_A\otimes B)(\phi_A\otimes A\otimes B)(A\otimes A\otimes\lambda)(A\otimes T\otimes B) \qquad (1_*)$$

and the following diagram (expressing the associativity of ϕ) is commutative:

$$\begin{tikzcd}
 & B\otimes A\otimes B \arrow[r, "T\otimes B"] & A\otimes B\otimes B \arrow[r, "A\otimes\lambda"] & A\otimes A\otimes B \arrow[dr, "\phi_A\otimes B"] & \\
B\otimes B\otimes B \arrow[ur, "B\otimes\lambda"] \arrow[dr, "\lambda\otimes B"'] & & & & C \quad (2_*) \\
 & A\otimes B\otimes B \arrow[rr, "A\otimes\lambda"] & & A\otimes A\otimes B \arrow[ur, "\phi_A\otimes B"'] &
\end{tikzcd}$$

Conversely, if $\lambda\colon B\otimes B\to C$ *is a map of G.C. modules satisfying* (2_*), $j_B\lambda=\phi_B$; *and if* ϕ *is defined by* (1_*), *then* $\lambda=\phi(i_B\otimes i_B)$ *and* (C,ϕ) *is a T.T.A.*

Remark. In the notation of 8.10_* this lemma merely expresses the fact that ϕ is known if $b\circ b'$ is given for $b,\ b'\in B$.

8.2^* LEMMA. *If* (C,ψ) *is a T.T.C. and* $\mu=(j_A\otimes j_A)\psi\colon C\to A\otimes A$, *then*

$$\psi=(A\otimes T\otimes B)(\mu\otimes B\otimes B)(A\otimes B\otimes\psi_B)(A\otimes\psi_B) \qquad \text{———}\ (1^*)$$

and the following diagram is commutative:

$$\begin{tikzcd}
 & A\otimes B\otimes A \arrow[dl, "\mu\otimes A"'] & A\otimes A\otimes B \arrow[l, "A\otimes T"'] & A\otimes B\otimes B \arrow[l, "\mu\otimes B"'] & \\
\otimes A\otimes A & & & & C \quad (2^*) \arrow[ul, "A\otimes\psi_B"'] \arrow[dl, "A\otimes\psi_B"] \\
 & A\otimes A\otimes B \arrow[ul, "A\otimes\mu"] & & A\otimes B\otimes B \arrow[ll, "\mu\otimes B"] &
\end{tikzcd}$$

Conversely, if $\mu\colon C\to A\otimes A$ *is a map of G.C. modules satisfying* (2^*), $\mu i_A=\psi_A$; *and if* ψ *is defined by* (1^*), *then* $\mu=(j_A\otimes j_A)\psi$ and (C,ψ) *is a T.T.C.*

In view of 8.2_*, 8.2^* a T.T.A. or T.T.C. is uniquely defined by its "*twisting function*" λ or μ respectively.

8.3_* Let $E=(A\otimes B,\phi)$ be a T.T.A. and let $a\colon A\to A'$ be a map of algebras. Let $\lambda\colon B\otimes B\to A\otimes B$ be the twisting functions, 8.2_*, for ϕ; define $\lambda'\colon B\otimes B\to A'\otimes B$ by $\lambda'=(a\otimes B)\lambda$. It is immediate that λ' satisfies the condition 8.2_* (2) for a twisting function; this defines a T.T.A. $(A'\otimes B,\phi')$ which can easily be shown to have the defining property of a_*E, cf. 5.4_*. The map $a_*\colon A\otimes B\to A'\otimes B$ of 5.4_* is $a\otimes B$.

Thus:

LEMMA. *If the T.T.A.* $(A\otimes B,\phi)$ *with twisting function* λ *represents the extension* E *and* $a\colon A\to A'$ *is a map of algebras, then* $(a\otimes B)\lambda$ *is a twisting function for* a_*E.

Note. We do not, in this theory, need to restrict ourselves to central a; cf. the remarks following 8.11_*.

8.3^* LEMMA. *If the T.T.C.* $(A\otimes B,\psi)$ *with twisting functions* μ *represents the extension* E *and* $b: B'\to B$ *is a map of co-algebras, then* $\mu(A\otimes b)$ *is a twisting function for* b^*E.

8.4 *Definition.* $(A\otimes B = C, \phi, \phi_A, \phi_B, \psi, \psi_A, \psi_B)$ of briefly (C,ϕ,ψ) will be called a twisted tensor product of Hopf-algebras (T. T. H.) if

(i) (C,ϕ,ψ), (A,ϕ_A,ψ_A), (B,ϕ_B,ψ_B) are Hopf-algebras.

(ii_*) (C,ϕ,ϕ_A,ϕ_B) is a T. T. A.

(ii^*) (C,ψ,ψ_A,ψ_B) is a T. T. C.

8.41 LEMMA. *If* ϕ_A, ψ_B *are commutative and* (C,ϕ,ψ) *is a T.T.H., then* $E(A,B)$ *is a central extension of Hopf-algebras.*

Conversely, given a central extension of Hopf-algebras, cf. 6.1, *if we use* Γ *to "transfer" the product and co-product of* C *to* $A\otimes B$, *we obtain a T.T.H.*

In this case we say that the T. T. H. is central.

8.5 LEMMA. *Let* (C,ϕ,ψ) *be a T.T.H. and let* λ, μ *be defined as in* 8.2_*, 8.2^*, *and let*

$$\alpha = j_A\lambda : B\otimes B\to A$$
$$\beta = \mu i_B : B\to A\otimes A$$

then

$$\lambda = (\alpha\otimes\phi_B)(B\otimes T\otimes B)(\psi_B\otimes\psi_B) \qquad \text{———} (3_*)$$
$$\mu = (\phi_A\otimes\phi_A)(A\otimes T\otimes A)(\psi_A\otimes\beta) \qquad \text{———} (3^*)$$

Proof. Notice that, now that we are in a self-dual situation, to every formula there corresponds a *specific* dual one; the word "reciprocal" might be used, as in the theory of a given conic section. 1_*, 2_*, 3_* are reciprocals of 1^*, 2^*, 3^*; in this reciprocation there correspond $A\leftrightarrow B$, $i\leftrightarrow j$, $\phi\leftrightarrow\psi$, $\lambda\leftrightarrow\mu$, $\alpha\leftrightarrow\beta$. We prove (3_*); the proof of (3^*) is "reciprocal."

First, we note

$$\begin{aligned}(j_A\otimes j_B)\psi &= (j_A\otimes C)(C\otimes j_B)\psi\\ &= (j_A\otimes C)(A\otimes\psi_B) \qquad \text{by } 8.1^* \quad (1)\\ &= (A\otimes\epsilon_B\otimes B)(A\otimes\psi_B)\\ &= A\otimes B\end{aligned}$$

so that

$$(j_A \otimes j_B)\psi = C \tag{4*}$$

Next, $$\begin{aligned}(A \otimes \psi_B) i_B &= (A \otimes \psi_B)(\eta_A \otimes B) = (\eta_A \otimes \psi_B)\\ &= (\eta_A \otimes B \otimes B)(K \otimes \psi_B)\\ &= (i_B \otimes B)\psi_B\end{aligned}$$

so that

$$(A \otimes \psi_B) i_B = (i_B \otimes B)\psi_B \tag{5*}$$

and we can compute:

$$\begin{aligned}\lambda &= \phi(i_B \otimes i_B)\\ &= (j_A \otimes j_B)\psi\phi(i_B \otimes i_B) && \text{by } (4^*)\\ &= (j_A \otimes j_B)(\phi \otimes \phi)(C \otimes T \otimes C)(\psi \otimes \psi)(i_B \otimes i_B)\\ & && \text{because } C \text{ is a Hopf-algebra}\\ &= (j_A\phi \otimes j_B\phi)(C \otimes T \otimes C)(\psi \otimes \psi)(i_B \otimes i_B)\\ &= (j_A\phi \otimes \phi_B)(C \otimes C \otimes j_B \otimes j_B)(C \otimes T \otimes C)(\psi \otimes \psi)(i_B \otimes i_B)\\ & && \text{by } 8.1_* \quad (3)\\ &= (j_A\phi \otimes \phi_B)(C \otimes T \otimes B)(C \otimes j_B \otimes C \otimes j_B)(\psi \otimes \psi)(i_B \otimes i_B)\\ &= (j_A\phi \otimes \phi_B)(C \otimes T \otimes B)(A \otimes \psi_B \otimes A \otimes \psi_B)(i_B \otimes i_B)\\ & && \text{by } 8.1^* \quad (1)\\ &= (j_A\phi \otimes \phi_B)(C \otimes T \otimes B)(i_B \otimes B \otimes i_B \otimes B)(\psi_B \otimes \psi_B)\\ & && \text{by } (5^*)\\ &= (j_A\phi \otimes \phi_B)(i_B \otimes i_B \otimes B \otimes B)(B \otimes T \otimes B)(\psi_B \otimes \psi_B)\\ &= (\alpha \otimes \phi_B)(B \otimes T \otimes B)(\psi_B \otimes \psi_B) && \text{Q.E.D.}\end{aligned}$$

Note, incidentally, that $(B \otimes T \otimes B)(\psi_B \otimes \psi_B) = \psi_{B \otimes B}$.

8.6 In view of 8.5, 8.2_*, 8.2^*, the T.T.H. (C, ϕ, ψ) is completely determined by the pair (α, β). These satisfy 3 conditions which express the associativity of ϕ, that of ψ and the Hopf condition for C.

The associativity of ϕ is expressed by (2_*), 8.2_*. Let us denote by $X_1, X_2 \colon B \otimes B \otimes B \to C$ the top and bottom of that diagram, so that 2_* becomes $X_1 = X_2$. Now, define

$$\bar{X}_k = j_A X_k \colon B \otimes B \otimes B \to A \qquad (k = 1, 2) \qquad (6_*)$$

Similarly, let $Y_1, Y_2 \colon C \to A \otimes A \otimes A$ denote the top and bottom of 2^* and define

$$\bar{Y}_k = Y_k i_B \colon B \to A \otimes A \otimes A \tag{6*}$$

8.61 Let us write

$$\phi_B{}^3 = \phi_B(B\otimes\phi_B) = \phi_B(\phi_B\otimes B) \text{ and } \psi_A{}^3 = (A\otimes\psi_A)\psi_A = (\psi_A\otimes A)\psi_A.$$

LEMMA.

$$X_k = (\bar{X}_k\otimes\phi_B{}^3)\psi_{B\otimes B\otimes B} \qquad (7_*)$$

$$Y_k = \phi_{A\otimes A\otimes A}(\psi_A{}^3\otimes\bar{Y}_k) \qquad (7^*)$$

From this, it follows that $X_1 = X_2$ if and only if $\bar{X}_1 = \bar{X}_2$, $Y_1 = Y_2$ if and only if $\bar{Y}_1 = \bar{Y}_2$. The explicit formulas for these homomorphisms are:

$$\bar{X}_1 = \phi_A(\alpha\otimes\alpha)(2,3,1,4)(B\otimes B\otimes B\otimes\phi_B)(1,2,4,3,5)(B\otimes\psi_B\otimes\psi_B) \qquad (8_*)$$

$$\bar{X}_2 = \phi_A(\alpha\otimes\alpha)(B\otimes B\otimes\phi_B\otimes B)(1,3,2,4,5)(\psi_B\otimes\psi_B\otimes B)$$

$$\bar{Y}_1 = (\phi_A\otimes\phi_A\otimes A)(1,3,2,4,5)(\psi_A\otimes A\otimes A\otimes A)(1,3,4,2)(\beta\otimes\beta)\psi_B \qquad (8^*)$$

$$\bar{Y}_2 = (A\otimes\phi_A\otimes\phi_A)(1,2,4,3,5)(A\otimes\psi_A\otimes A\otimes A)(\beta\otimes\beta)\psi_B.$$

Next, we deal with the Hopf condition for C. As it stands, it asserts the equality of 2 homomorphisms $A\otimes B\otimes A\otimes B \to A\otimes B\otimes A\otimes B$; by a trivial transformation and substituting from 8.2_*, 1_* and 8.2^*, 1^*, the Hopf-condition is the commutativity of the following diagram:

$$\begin{tikzcd}
A\otimes A\otimes B\otimes B \arrow[r, "\phi_A\otimes\lambda"] \arrow[d, "A\otimes A\otimes\psi_B\otimes\psi_B"] & A\otimes A\otimes B \arrow[r, "\phi_A\otimes B"] & A\otimes B \arrow[r, "A\otimes\psi_B"] & A\otimes B\otimes B \arrow[r, "\mu\otimes\psi_B"] & A\otimes A\otimes B\otimes \\
A\otimes A\otimes B\otimes B\otimes B\otimes B \arrow[d, "{(1,3,4,2,5,6)}"] & & & & A\otimes A\otimes A\otimes A\otimes B\otimes \arrow[u, swap, "\phi_A\otimes\phi_A\otimes B\otimes B"] \\
A\otimes B\otimes B\otimes A\otimes B\otimes B \arrow[d, "\mu\otimes\psi_B\otimes\mu\otimes\psi_B"] & & & & A\otimes A\otimes B\otimes A\otimes A\otimes \arrow[u, swap, "{(1,2,4,5,3,6)}"] \\
A\otimes A\otimes B\otimes B\otimes A\otimes A\otimes B\otimes B \arrow[rrrr, "{(1,5,3,7,2,6,4,8)}"] & & & & A\otimes A\otimes B\otimes B\otimes A\otimes A\otimes B\otimes B \arrow[u, swap, "\phi_A\otimes\lambda\otimes\phi_A\otimes\lambda"]
\end{tikzcd}$$

(9)

Let us denote by Z_1 the top composition in (9), and by Z_2 the composition of the other 3 sides, so that the Hopf-condition is $Z_1 = Z_2$. We define

$$\bar{Z}_k = (j_A\otimes j_A)(1,3,2,4)Z_k(1,3,2,4)(i_B\otimes i_B) \qquad (k=1,2) \qquad (10)$$

and prove

LEMMA. *Let ϕ_A, ψ_B be commutative. Then:*

$$Z_k = (\phi^3{}_{A\otimes A} \otimes B \otimes B)(\psi_A \otimes \psi_A \otimes \bar{Z}_k \otimes \phi_B \otimes \phi_B)(A \otimes A \otimes \psi^3{}_{B\otimes B}) \tag{11}$$

Hence $\bar{Z}_1 = \bar{Z}_2$ implies $Z_1 = Z_2$.

We notice the following explicit formulas:

$$\begin{aligned}\bar{Z}_1 &= \phi_{A\otimes A}(\psi_A \otimes \beta)(\alpha \otimes \phi_B)\psi_{B\otimes B}\\ \bar{Z}_2 &= (\phi_A \otimes \phi_A)(\phi_A \otimes \alpha \otimes \phi_A \otimes \alpha)(1,5,3,7,2,6,4,8)(\beta \otimes \psi_B \otimes \beta \otimes \psi_B)(\psi_B \otimes \psi_B)\end{aligned} \tag{12}$$

In order to express conveniently what we have proved we adopt the following

8.62 *Definition.* (α, β) will be called an *admissible (A,B)-pair* if $\alpha\colon B \otimes B \to A$, $\beta\colon B \to A \otimes A$ are homomorphisms of G. C. modules such that $\bar{X}_1 = \bar{X}_2$, $\bar{Y}_1 = \bar{Y}_2$, $\bar{Z}_1 = \bar{Z}_2$ where $\bar{X}_k$, $\bar{Y}_k$, $\bar{Z}_k$ are given by (8_*), (8^*), (12).

8.63 THEOREM. *The formulas*

$$\begin{aligned}\alpha &= j_A\phi(i_B \otimes i_B)\\ \beta &= (j_A \otimes j_A)\psi i_B.\end{aligned}$$

associate an admissible (A,B)-pair (α,β) with every T.T.H. $(A \otimes B, \phi, \psi)$.

Conversely, if (α,β) is an admissible (A,B)-pair and ϕ_A, ψ_B are commutative then the formulas $8.2_*(1_*)$, $8.5(3_*)$, $8.2^*(1^*)$, $8.6(3^*)$ *define a central T.T.H. $(A \otimes B, \phi, \psi)$.*

8.7 Let (α,β) be an admissible (A,B)-pair and let $a\colon A \to A'$, $b\colon B' \to B$ be maps of Hopf algebras. Then it is easily seen that, defining

$$\begin{aligned}\alpha' &= a\alpha(b \otimes b)\\ \beta' &= (a \otimes a)\beta b\end{aligned}$$

(α', β') is an admissible (A', B')-pair.

If (α,β) represents the central extension E, then it is clear from 8.3_*, 8.3^*, that (α',β') represents $a_*b^*E = b^*a_*E$.

The inner reason for the existence of both kinds of induced structure for central extensions of Hopf-algebras is now clear: Such an extension can be defined in terms of functions, and identities between such functions, all of the form

(covariant functor of B) → (covariant functor of A).

8.8 We now introduce the notions that will allow us to prove Theorem 6.61. If Y, X are G. C. modules, by $G(Y,X)$ let us denote the set of all maps of G. C. modules $Y \rightarrow X$.

If (Y,ψ) is a co-algebra and (X,ϕ) an algebra, we introduce a binary operation $\smile$, the "cup-product," into $G(Y,X)$, as follows:

If $f, g \in G(Y,X)$, we define

$$f \smile g = \phi(f \otimes g)\psi$$

It is easily proved that the associativity of ϕ, ψ implies that of $\smile$; and that the map $1 = \eta_X \epsilon_Y$ is the unit for $\smile$. If ϕ, ψ are both commutative, then so is $\smile$.

8.81 LEMMA. *If Y is a co-algebra and X an algebra, then, with the binary operation $\smile$, $G(Y,X)$ is a group*; 1 *is the unit.*

For a proof, see [1], 6.8.

8.9 Now, let the extension classes x, y be represented by the T. T. H. $E_t = (A \otimes B, \phi_t, \psi_t)$ $(t = x, y)$ with admissible (A,B)-pairs (α_t, β_t). $E_x \otimes E_y$ is the $(A \otimes A, B \otimes B)$ extension

$$A \otimes A \xrightarrow{i_A \otimes i_A} C \otimes C \xrightarrow{j_B \otimes j_B} B \otimes B$$

which is easily seen to be equivalent to the T. T. H.

$$A \otimes A \xrightarrow{i_{A \otimes A}} A \otimes A \otimes B \otimes B \xrightarrow{j_{B \otimes B}} B \otimes B.$$

Let (α_z, β_z) be the admissible $(A \otimes A, B \otimes B)$-pair of this T. T. H. The product for $C \otimes C$ *is* $(\phi_x \otimes \phi_y)(C \otimes T \otimes C)$. The corresponding product for $A \otimes A \otimes B \otimes B$ is

$$\phi_z = P(\phi_x \otimes \phi_y)(C \otimes T \otimes C)(P \otimes P)$$

where we have written $P = A \otimes T \otimes B : A \otimes A \otimes B \otimes B \rightarrow A \otimes B \otimes A \otimes B$, or its inverse. Notice that $i_{A \otimes A} = P(i_A \otimes i_A)$ and $j_{B \otimes B} = (j_B \otimes j_B)P$.

Now:

$$\begin{aligned}
\alpha_z &= j_{A \otimes A}\, \phi_z(i_{B \otimes B} \otimes i_{B \otimes B}) \\
&= (j_A \otimes j_A)PP(\phi_x \otimes \phi_y)(C \otimes T \otimes C)(P \otimes P)(P \otimes P)(i_B \otimes i_B \otimes i_B \otimes i_B) \\
&= (j_A \otimes j_A)(\phi_x \otimes \phi_y)(C \otimes T \otimes C)(i_B \otimes i_B \otimes i_B \otimes i_B) \\
&= (j_A \otimes j_A)(\phi_x \otimes \phi_y)(i_B \otimes i_B \otimes i_B \otimes i_B)(B \otimes T \otimes B) \\
&= (\alpha_x \otimes \alpha_y)(B \otimes T \otimes B).
\end{aligned}$$

By 6. 5, $x+y$ is represented by $\phi_{A*}\psi_B^*E_z$; let (α,β) be the corresponding admissible (A,B)-pair. By 8. 7.

$$\begin{aligned}\alpha &= \phi_A\alpha_z(\psi_B\otimes\psi_B)\\ &= \phi_A(\alpha_x\otimes\alpha_y)(B\otimes T\otimes B)(\psi_B\otimes\psi_B)\\ &= \phi_A(\alpha_x\otimes\alpha_y)\psi_{B\otimes B}\\ &= \alpha_x\cup\alpha_y.\end{aligned}$$

We have proved

8. 91 LEMMA. *If the admissible* (A,B)-*pairs* (α_x,β_x), (α_y,β_y) represent *the extension classes* x, y *respectively, then* $x+y$ (cf. 6. 5) *is represented by* $(\alpha_x\cup\alpha_y,\beta_x\cup\beta_y)$; *in particular, this last pair is admissible.*

It is easily seen that $(1,1)$ is the admissible pair for the "trivial" T. T. H. —and thus the assertion concerning the unit in 6. 6 is proved.

8. 92 It remains to establish the existence of inverses. It is clear from what has been said that all we have to prove is

LEMMA. *Let* $\alpha\in G(B\otimes B,A)$ *and* $\beta\in G(B,A\otimes A)$ *be such that* (α,β) *is an admissible pair; let* ψ_B *and* ϕ_A *be commutative; then* (α^{-1},β^{-1}) *is an admissible pair.*

Proof. Let us write $\bar{X}_1(\alpha)$, $\bar{Z}_1(\alpha,\beta)$, etc., for the functions $\bar{X}_1$, $\bar{Z}_1$, etc., of (α,β).

Now, observe that

$$\begin{aligned}&\bar{X}_k\in G(B\otimes B\otimes B,A)\\ &\bar{Y}_k\in G(B,A\otimes A\otimes A)\\ &\bar{Z}_k\in G(B\otimes B,A\otimes A).\end{aligned}$$

We are given

$$\bar{X}_1(\alpha)=\bar{X}_2(\alpha),\quad \bar{Y}_1(\beta)=\bar{Y}_2(\beta),\quad \bar{Z}(\alpha,\beta)=\bar{Z}_2(\alpha,\beta)$$

and we want to prove the corresponding identities for (α^{-1},β^{-1}); we shall clearly be done if we can prove

$$\begin{aligned}\bar{X}_k(\alpha^{-1})&=[\bar{X}_k(\alpha)]^{-1}\\ \bar{Y}_k(\beta^{-1})&=[\bar{Y}_k(\beta)]^{-1}\\ \bar{Z}_k(\alpha^{-1},\beta^{-1})&=[\bar{Z}_k(\alpha,\beta)]^{-1}\end{aligned}$$

and these formulas can, in fact, be proved: By a direct examination of the diagrams one proves

$$\bar{X}_k(\alpha^{-1})\cup\bar{X}_k(\alpha)=1 \text{ etc.}$$

These proofs are very complicated in the case of $\bar{X}_k$, $\bar{Y}_k$, but much simpler for $\bar{Z}_k$. We give the easiest case —$\bar{Z}_1$:

$$\begin{aligned}\bar{Z}_1(\alpha,\beta) &= \phi_{A\otimes A}(\psi_A\otimes\beta)(\alpha\otimes\phi_B)\psi_{B\otimes B}\\ &= \phi_{A\otimes A}(\psi_A\alpha\otimes\beta\phi_B)\psi_{B\otimes B}\\ &= \psi_A\alpha\cup\beta\phi_B.\end{aligned}$$

Due to the commutativity and associativity of $\cup$, all we have to prove is the reciprocal formulas

$$\psi_A\alpha^{-1}=(\psi_A\alpha)^{-1},\quad \beta^{-1}\phi_B=(\beta\phi_B)^{-1}.$$

Now:

$$\begin{aligned}\psi_A\alpha^{-1}\cup\psi_A\alpha &= \phi_{A\otimes A}(\psi_A\alpha^{-1}\otimes\psi_A\alpha)\,\psi_{B\otimes B}\\ &= \phi_{A\otimes A}(\psi_A\otimes\psi_A)(\alpha^{-1}\otimes\alpha)\,\psi_{B\otimes B}\\ &= (\phi_A\otimes\phi_A)(A\otimes T\otimes A)(\psi_A\otimes\psi_A)(\alpha^{-1}\otimes\alpha)\psi_{B\otimes B}\\ &= \psi_A\phi_A(\alpha^{-1}\otimes\alpha)\psi_{B\otimes B}\\ &\qquad\qquad \text{by the Hopf-condition for } A\\ &= \psi_A\eta_A\epsilon_{B\otimes B}=\eta_{A\otimes A}\,\epsilon_{B\otimes B}\\ &= 1. \qquad\qquad \text{Q.E.D.}\end{aligned}$$

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] John W. Milnor and John C. Moore, *On the structure of Hopf Algebras* (Mimeographed, Princeton).

[2] J. F. Adams, " On the non-existence of elements of Hopf invariant one," *Annals of Mathematics*, vol. 72 (1960).

[3] E. H. Brown, " Twisted tensor products I," *Annals of Mathematics*, vol. 67 (1958).

[4] V. K. A. M. Gugenheim, *On twisted Tensor Products of Algebras* (Mimeographed, Johns Hopkins).

[5] *Seminaire Henri Cartan*, 12e année (1959/60), Fascicule 1 (2e édition) (Paris). (Exp. 7, J. C. Moore).

REMARKS ON THE PAPER "GENERALIZATION OF A THEOREM OF MANDELBROJT."*

By James A. Jenkins.

1. Shortly after the appearance of the paper [2] I. I. Hirschman pointed out to me that the inequality in the display at the bottom of page 809 actually should be in the opposite sense so that the principle of extension of domain does not lead to the result asserted. However, as I observed to him at that time, this does not affect the essential idea of the paper since this principle was applied chiefly as a technical device to avoid explicit evaluation of the harmonic measure associated with the domain there denoted by $D(\epsilon)$. As an alternative one can first carry out the limit process corresponding to that at the top of page 810 then insert the actual value of the element of harmonic measure with respect to the semi-circle $|z| < r$, $\Im z > 0$. The remaining arguments then show that the statement of Theorem 1 remains true if we replace the expression there denoted by $U(r)$ by

$$(1)\qquad \pi^{-1}\int_{-r}^{r}(r^2-1)(r^2-x^2)(r^4+x^2)^{-1}(1+x^2)^{-1}\log|\phi(x)|\,dx$$

While this statement is not quite as strong as the previous assertion it suffices for the derivation of the same results obtained before as consequences, namely results of Hardy, Morgan, Ingham and Paley and Wiener. While the expression (1) may seem somewhat complicated it is easily seen that we can by surrendering some sharpness use an expression differing from $U(r)$ only by the introduction of constant factors, for example

$$(\tfrac{3}{4}-\epsilon)\pi^{-1}\int_{-\frac{1}{2}r}^{\frac{1}{2}r}(1+x^2)^{-1}\log|\phi(x)|\,dx$$

for arbitrary positive ϵ.

2. On the other hand it is clear that a semi-circle may not be the most advantageous comparison domain. If we confine ourselves to domains based on the segment $[-r, r]$ and lying in the strip $0 < \Im z < r$ we see at once that for any such domain the element of harmonic measure on $[-r, r]$ will

---

* Received March 21, 1962.

be majorized by that relative to the strip so that the latter would be the most advantageous comparison domain. However it cannot be used directly since there is no bound available for $\log|F(z)|$ on the portion of real axis complementary to the segment $[-r, r]$. In order to use the strip Hirschman [1] placed very strong restrictions on the rate of growth of the functions treated. By a modification of our previous argument it can be shown that it is possible to obtain essentially the full effect of using the strip without placing any restriction on the functions in question beyond that given in [2]. The result so obtained is in a certain sense best possible. Of course there is no inherent reason that the half-length of the base should be equal to the height of the strip and extensions are easily obtained in this direction.

THEOREM 1. *Let* $f(t)$ *and* $\phi(x)$ *be in* $L^2(-\infty,\infty)$ *and be transforms of each other in the sense of Plancherel's Theorem,* i. e.,

$$\phi(x) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} e^{-ixt} f(t)\,dt, \quad f(t) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} e^{itx}\phi(x)\,dx$$

where equality is understood in the sense of the mean square limit. Let the element of harmonic measure at the point $\xi + i\eta$ *in the strip* $0 < \Im z < r$ *with respect to the strip for the boundary element at the boundary point* $x + iy$ *be denoted by* $k(\xi,\eta;x,y;r)\,ds$ *(*ds *element of length). Let*

$$U(r) = \int_{-r}^{r} k(0,1;x,0;r)\log|\phi(x)|\,dx. \tag{2}$$

Let there exist a function $V(y)$*, non-decreasing for* $y > 0$ *such that*

$$\int_{V(y)}^{\infty} e^{ty}|f(t)|\,dt = O(e^{yV(y)}) \quad as\ y\to\infty. \tag{3}$$

Then if for some $\delta > 0$

$$\varliminf_{r\to\infty}((1+\delta)V(r) + U(r)) = -\infty \tag{4}$$

f *and* ϕ *are both almost everywhere zero.*

As in [2] there is a corresponding application to quasianalytic functions for which we use the same notation as in [2; p. 810].

THEOREM 2. *Let* $g(t)$ *belong to the class* $C\{M_n\}$ *where* $\varliminf_{n\to\infty}(M_n)^{1/n} = \infty$ *and satisfy*

$$\int_{V(y)}^{\infty} e^{ty}|g(t)|\,dt = O(e^{yV(y)}), \qquad y\to\infty$$

with $V(y)$ *non-decreasing for* $y > 0$. *Let*

$$W(r,h) = 2\int_0^r k(0,1;x,0;r)\log T(x/h)\,dx$$

satisfy for some $h > k$ *and* $\delta > 0$

$$\varliminf_{r\to\infty}((1+\delta)V(r) - W(r,h)) = -\infty.$$

Then $g(t) \equiv 0$.

REFERENCES.

[1] I. I. Hirschman, Jr., "On the behavior of Fourier transforms at infinity and on quasi-analytic classes of functions," *American Journal of Mathematics*, vol. 72 (1950), pp. 200-213.

[2] James A. Jenkins, "Generalization of a theorem of Mandelbrojt," *American Journal of Mathematics*, vol. 73 (1951), pp. 807-812.

ERRATA.

"Representation of a semigroup by transformations acting transitively on a set," by E. J. Tully, Jr., *American Journal of Mathematics*, vol. 83 (1961), pp. 533-541.

	for	read
p 539, line 9 from bottom:	transitive	not transitive
p 539, line 8 from bottom:	rows	columns
	not left	right
p. 541, line 13:	rows	columns
p. 541, line 14:	$p_{ik}a = p_{jk}b$	$p_{ki}a = p_{kj}b$
	rows	columns
p. 541, line 18:	rows	columns

ON THE STRUCTURE OF MANIFOLDS.*

By S. SMALE.[1]

In this paper, we prove a number of theorems which give some insight into the structure of differentiable manifolds.

The methods, results and some notation of [13], hereafter referred to as GPC, and [12] will be used. These two papers and [14] can be considered as a starting point for this one. The main theorems in these papers are special cases of the theorems here.

Among the most important theorems in this paper are 1.1 and 6.1.

Some conversations with A. Haefliger were helpful in the preparation of parts of this paper.

Everything will be considered from the differentable, equivalently C^∞, point of view; manifolds, imbeddings, and isotopes will be C^∞.

Section 1. We give a necessary and sufficient condition for two closed simply connected manifolds of dimension greater than four to be diffeomorphic. The condition is *h-cobordant,* first defined by Thom [16] for the combinatorial case, and developed by Milnor [9], and Kervaire and Milnor [7] for the differentiable case (sometimes previously h-cobordant has been called J-equivalent). It involves a combination of homotopy theory and cobordism theory. More precisely, two closed connected oriented manifolds $M_1{}^n$, $M_2{}^n$ are h-cobordant if there exists an oriented compact manifold W with ∂W (the boundary of W) diffeomorphic to the disjoint union of M_1 and $-M_2$, and each component of ∂W is a deformation retract of W.

THEOREM 1.1. *If $n \geqq 5$, and two closed oriented simply connected manifolds $M_1{}^n$ and $M_2{}^n$ are h-cobordant, then M_1 and M_2 are diffeomorphic by an orientation preserving diffeomorphism.*

It has been asked by Milnor whether h-cobordant manifolds in general are diffeomorphic, problem 5, [9]. Subsequently, Milnor himself has given a counter-example of 7-dimensional manifolds with fundamental group Z_7, h-cobordant but not diffeomorphic [10]. Thus the condition of simple-connectedness is necessary in Theorem 1.1.

---

* Received July 18, 1961.

[1] The author is an Alfred P. Sloan Fellow.

Theorem 1.1 was proved in special cases in [13] and [14]. These special cases were applied to show that every sphere not of dimension four or six has a finite number of differentiable structures. The six dimensional case is taken care of by the following.

COROLLARY 1.2. *Every homotopy 6-sphere is diffeomorphic to* S^6.

This follows from 1.1 and the result of Kervaire and Milnor [7] that every homotopy 6-sphere is h-cobordant to S^6.

COROLLARY 1.3. *The semigroup of 2-connected closed 6-manifolds is generated by* $S^3 \times S^3$.

This follows from 1.2 and [15].

Haefliger [2] has extended the notion of h-cobordant to the relative case. Let V_1, V_2, M_1, M_2 be closed oriented, connected manifolds with $V_i \subset M_i$, $i = 1, 2$. According to Haefliger (M_1, V_1), (M_2, V_2) are *h-cobordant* if there is a pair (M, V) (i.e., $V \subset M$) with $\partial M = M_1 - M_2$, $\partial V = V_1 - V_2$ and $M_i \to M$, $V_i \to V$ homotopy equivalences. Then 1.1 can be extended to the relative case.

THEOREM 1.4. *Suppose* (M_1^n, V_1^k) *and* (M_2^n, V_2^k) *are h-cobordant,* $k \geqq 5$, $\pi_1(V_i) = \pi_1(M_i - V_i) = 1$. *Then there is an orientation preserving diffeomorphism of* M_1 *onto* M_2 *sending* V_1 *onto* V_2.

By taking V_i empty (the proof of 1.4 is valid for this case also), one can consider 1.1 as a special case of 1.4.

Actually we obtain much stronger theorems which will imply 1.4. The proof of 1.4 is completed in Section 3.

It would not be surprising if the hypothesis of simple connectedness in these theorems could be weakened using torsion invariants (see [10], for example).

Theorem 1.4 has application in the theory of knots except in codimension two.

Section 2. The main theorem we prove in this section is the following. Here we use the notation of GPC.

THEOREM 2.1. *Let* M^n *be a compact manifold with a simply connected boundary component* Q. *Let* $V = \chi(M, Q; f; m)$ *where* $f: \partial D_0^m \times D_0^{n-m} \to Q$ *is an embedding,* $m > 2$, $n - m > 3$. *Suppose* $W = \chi(V, Q_1; g_1, \cdots, g_r; m+1)$ *where* Q_1 *is the component of* ∂V *corresponding to* Q *and suppose that* $H_m(W, M)$ *is zero. Then* W *is of the form* $\chi(M, Q; g'_1, \cdots, g'_{n-1}; m+1)$.

Note that an example of Mazur [8] shows that the dimensional restriction is necessary here.

For the proof we use several lemmas.

LEMMA 2.2. *Let* M^n *be a compact manifold,* Q *a component of* ∂M, $n-m>1$, $V=\chi(M,Q;f;m)$, $W=\chi(V,Q_1;g_1,\cdots,g_r;m+1)$ *where* Q_1 *is the component of* ∂V *corresponding to* Q *and* $f\colon \partial D_0^m\times D_0^{n-m}\to Q$, $g_i\colon \partial D_i^{m+1}\times D_i^{n-m-1}\to Q$, *are imbeddings. Let* $F=q\times D_0^{n-m}\subset V$ *with* $q\notin \partial D_0^m$, $\partial F\subset\partial V$. *Suppose* ∂F *does not intersect* $g_i(\partial D_i^{m+1}\times 0)$, $i=1,\cdots,r-1$ and $g_r(\partial D_r^{m+1}\times 0)$ *intersects* ∂F *transversally in a single point. Then* W *is of the form*

$$\chi(M,Q;g'_1,\cdots,g'_{r-1};m+1).$$

Proof 2.2. In the proof of 2.2, we use without further mention, the fact that the diffeomorphism type of an n-manifold is not changed when an n-disk is adjoined by identifying an $(n-1)$ disk on the boundary of each under a diffeomorphism. See GPC, 3.4, J. Milnor, "Sommes de varietes différentiables et structures différentiables des sphères," *Bulletin de la Societe Mathematique de France,* vol. 87 (1959), pp. 439-444 and R. Palais, "Extending diffeomorphism," *Proceedings of the American Mathematical Society,* vol. 11 (1960), pp. 274-277.

We may assume, using the uniqueness of tubular neighborhoods that ∂F does not intersect $g_i(\partial D_i^{m+1}\times D_i^{n-m-1})$, $i=1,\cdots,r-1$.

Since $g_r(\partial D^{m+1}\times 0)$ is transversal to ∂F in ∂V, there exists a disk neighborhood L of $o=g_r(\partial D^{m+1}\times 0)\cap F$, $L=A^m\times B^{n-m-1}$, where $A^m\times 0$ is a disk neighborhood of o in $g_r(\partial D^{m+1}\times 0)$, $0\times B^{n-m-1}$ a disk neighborhood of o in ∂F, with (o,o) corresponding to o.

Now there exists a disk neighborhood D_a^m of the point $F\cap(D_0^m\times 0)$ in $D_0^m\times 0$ so small that if $N=D_a^m\times D_0^{n-m}\subset V$, then

(1) $N\cap \text{image}\, g_i=\emptyset$, $i=1,\cdots,r-1$, and

(2) $N\cap\text{image}\, g_r\subset L$.

Since both $D_a^m\times o$ and $A^m\times o$ (i.e. $A^m\times 0$) are transversal to ∂F in ∂V, we may assume using a diffeomorphism of V, and restricting L, that $A^m\times o$, $D_a^m\times o$ coincide, and that L coincides with image $g_r\cap N$.

The following statements are made under the assumptions that corners are smoothed via "straightening the angle," Section 1 of GPC or better [9]. Let $K=N\cup D_r^{n-m-1}\subset W$.

We claim that $K \cap Cl(W - K)$ is diffeomorphic to an $(n-1)$-disk. First $K \cap Cl(W - K)$ is $\partial D_a^m \times (\partial D_r^{m+1} \times D_r^{n-m-1})$ — interior L or

$$\partial D_a^m \times D_0^{n-m} \cup D_b^m \times D_r^{n-m-1}$$

where D_b^m is ∂D_r^{m+1} minus the interior of an m-disk. Furthermore $K \cap Cl(W - K)$ may be described as $\partial D_a^m \times D_0^{n-m}$ with $D_b^m \times D_r^{n-m-1}$ attached by an embedding $h\colon \partial D_b^m \times D_r^{n-m-1} \to \partial D_a^m \times D_0^{n-m}$ with the property that $h(\partial D_b^m \times 0)$ coincides with $\partial D_a^m \times c$ for some point $c \in D_0^{n-m}$. In fact, h is the restriction of g_r. This is the situation in the proof of 3.3 of GPC, where it was shown that the resulting manifold was a disk. Thus $K \cap Cl(W - K)$ is indeed an $(n-1)$-disk.

Since K is an n-disk, $K \cap Cl(W - K)$ an $(n-1)$-disk, we have that W is diffeomorphic to $Cl(W - K)$. On the other hand it is clear from the previous considerations that $Cl(W - K)$ is of the form

$$\chi(M, Q; g'_1, \cdots, g'_{r-1}, m+1).$$

This proves 2.2.

The next lemma follows from the method of Whitney [18] of removing isolated intersection points. The paper of A. Shapiro [11] makes this apparent (apply 6.7, 6.10, 7.1 of [11]).

LEMMA 2.3. *Suppose N^{n-m} is a closed submanifold of the closed manifold X^n and $f\colon M^m \to X$ is an imbedding of a closed manifold. Suppose also that M, N are connected, X is simply connected, $n - m > 2$, $m > 2$ and $b = f(M) \circ N$ is the intersection number of $f(M)$ and N. Then there exists an imbedding $f'\colon M \to X$ isotopic to f such that $f'(M)$ intersects N in b points, each with transversal intersection.*

LEMMA 2.4. *Let F_0^{n-m-1} be a submanifold of Q where Q is a component of the boundary of a compact manifold V^n, $n - m > 2$. Let $W = \chi(V, Q; g; m+1)$ where $g\colon \partial D_0^{m+1} \times D_0^{n-m-1} \to Q$ is an imbedding with b the intersection number $g(\partial D_0^{m+1} \times 0) \circ F_0$. For an imbedding h: $S^m \to Q \cap \partial W$, there is an imbedding $h'\colon S^m \to Q \cap \partial W$, isotopic to h in ∂W with $h'(S^m) \circ F_0 = h(S^m) \circ F_0 \pm b$, sign prescribed.*

Proof. Let D be the closed upper hemisphere of S^m, $\chi_0 \in \partial D_0^{n-m-1}$ and H^+, H^- be the closed upper, lower hemisphere respectively of $g(\partial D_0^{m+1} \times \chi_0)$. Then h is isotopic in $\partial W \cap Q$ to an imbedding $\bar{h}'\colon S^m \to \partial W \cap Q$, with $\bar{h}'(S^m) \cap g(\partial D_0^{m+1} \times \chi_0)$ equal H^+ with the orientation determined by the $\pm b$ of 2.4. This follows essentially from R. Palais, Extending Diffeomorphism, Proc. AMS, vol. 11 (1960), pp. 274-277, Theorem B, Corollary 1.

Next let $\bar{h}$ be $\bar{h}'$ followed by the reflection map $H^+ \to H^-$, so that $\bar{h}$, $\bar{h}'$: $D \to \partial W$ are naturally topologically isotopic. However $\bar{h}'$ is an angle on ∂D. By the familiar process of "straightening the angle" we modify $\bar{h}'$: $S^m \to \partial W \cap Q$ to an embedding h': $S^m \to \partial W \cap Q$. Our construction makes it clear that h' and h are isotopic in ∂W, and that h' has the desired property of 2.4.

We now prove 2.1. Let F be as in 2.2 and b_i be the algebrais intersection number $g_i(\partial D^{m+1} \times 0) \circ \partial F$, $i = 1, \cdots, r$. We first note that the b_i are relatively prime. This in fact follows from the homology hypothesis of the theorem.

The proof proceeds by induction on $\sum_{i=1}^{r} | b_i |$ and is started by 2.3 and 2.2. Suppose 2.1 is true in case $\sum_{i=1}^{r} | b_i |$ is $p - 1 > 0$.

We can say from the homotopy structure of W that $H_m(W, M)$ is $H_m(V, M)$ wtih the added relations $[\partial D_i^{m+1}] = 0$, $i = 1, \cdots, r$, where $[\partial D_i^{m+1}] \subset H_m(V, M) = Z$ and $H_m(V, M)$ is generated by $(D_0^m, \partial D_0^m)$.

Since $H_m(W, M) = 0$, $[\partial D_i^{m+1}]$ are relatively prime. On the other hand, since $D_0^m \times 0 \circ F = 1$, we have that $[\partial D_i^{m+1}] = b_i$. So the b_i, $i = 1, \cdots, r$ are relatively prime.

Since the b_i are relatively prime, there exists, i_0, i_1, $i_0 \neq i_1$ with $| b_{i_0} | \geqq | b_{i_1} | > 0$. One now applies 2.4 to reduce $| b_{i_0} |$ by $| b_{i_1} |$ using the covering homotopy property as in Section 2 of GPC. The induction hypothesis applies and we have proved 2.1.

LEMMA 2.5. *Let* $n \geqq 2m + 1$, $(n, m) \neq (4,1)$, (3.1), (5.2), (7.3), M^n *be a compact manifold with a simply connected boundary component* Q *and* $V = \chi(M, Q; f; m)$ *where* f: $\partial D^m \times D^{n-m} \to Q$ *is a contractible imbedding. Let* Q_1 *be the component of* ∂V *corresponding to* Q *and* $W = \chi(V, Q_1; g; m+1)$ *where* g: $\partial D_1^{m+1} \times D_1^{n-m} \to Q_1$. *Then if the homomorphism* $\pi_m(V, M) \to \pi_m(W, M)$ *induced by inclusion is zero,* W *is diffeomorphic to* M.

We use the following for the proof of 2.5.

LEMMA 2.6. *Let* Y *be a simply connected polyhedron and* Z *an* $(m-1)$-*connected polyhedron. Then* $\pi_m(Y \vee Z) = \pi_m(Y) + \pi_m(Z)$.

This is a standard fact in homotopy theory. For example it follows from [6], V.3.1 and the relative Hurewicz theorem.

Using 2.6 it follows easily that $\pi_m(Q_1) = \pi_m(Q) + \pi_m(S^m)$.

Then from the homotopy hypothesis it follows that the homotopy class γ of g restricted to $\partial D_1^{m+1} \times 0$ is of the form $a + g_1$ where $a \in \pi_m(Q)$ and g_1 generates $\pi_m(S^m)$.

Since P is contractible, $V = M + H$, where H is an $(n-m)$-cell bundle over S^m, and also $Q_1 = Q + \partial H$. Then let $g'_1 \colon \partial D^{m+1} \to Q$ be an imbedding representing a and $g'_2 \colon \partial D^{m+1} \to \partial H$ an imbedding intersecting ∂F transversally in a single point where F is the same as n 2.2. Then by the sum construction we obtain $g' \colon \partial D^{m+1} \times 0 \to Q_1$ realizing γ with the property that $g'(\partial D_1{}^{m+1} \times 0)$ intersects ∂F transversally in a single point where F is the same as in 2.2. Application of 2.4 of GPC and 2.2 finishes the proof.

Section 3. Among other things, we apply the theory of Section 2 to obtain 1.4.

THEOREM 3.1. *Let W^n be a manifold (not necessarily compact), $n > 5$, with ∂W the disjoint union of simply-connected manifolds M_1 and M_2 where the inclusion $M_i \to W$ are homotopy equivalences. Suppose $j \colon V_0 \to M_1$ is the inclusion of a compact manifold V_0 into M which is a homotopy equivalence and there is an imbedding $\alpha \colon Cl(M_1 - V_0) \times [1,2] \to W$ such that* a) *the complement of the image of α has compact closure and* b)

$$\alpha(Cl(M_1 - V_j) \times n) \subset M_i, \qquad i = 1, 2,$$

α restricted to $Cl(M_1 - V_0) \times 1$ is j. Then α can be extended to a diffeomorphism $M_1 \times [1,2] \to W$.

Proof of 3.1. *Let* $I_0 = [-\tfrac{1}{2}, n + \tfrac{1}{2}]$ and replace $[1,2]$ in the statement of 3.1 by I_0, denoting the projection $Cl(M_1 - V_0) \subset I_0 \to I_0$ by f_0. We may suppose that points under α have been identified so that $Cl(M_1 - V_0) \times I_0 \subset W$. Then by the results of [12] one can find a non-degenerate C^∞ real function f on W such that a) f restricted to $Cl(M_1 - V_0) \times I_0$ is f_0, b) at a critical point the value of f is the index and c) $f(M_1) = -\tfrac{1}{2}$, $f(M) = n + \tfrac{1}{2}$.

Let $X_p = f^{-1}[-\tfrac{1}{2}, p + \tfrac{1}{2}]$. We will show inductively that by suitable modifications of f which also satisfy a), b) and c), we can assume X_p is a product $M_1 \times I$ (or equivalently the modified f has no critical points of index $\leqq p$).

First by 5.1 of GPC, note that we may assume that the function f has no critical points of index 0. Next by the method in Section 7 of GPC, using the fact that $\pi_1(M_1) = \pi_1(W) = 1$, we can similarly assume that there are no critical points of f of index 1.

We are not quite yet in the dimension range where 2.1 applies, but we apply 2.5 to eliminate a critical point of f of index 2 if it occurs, as follows.

We have that $X_2 = \chi(X_1, Q_1; f_1, \cdots, f_k; 2)$, $X_3 = \chi(X_2, Q_2; g_1, \cdots, g_r; 3)$ where $Q_1 = f^{-1}(1\tfrac{1}{2})$, $Q_2 = f^{-1}(2\tfrac{1}{2})$. It follows from the homotopy hypothesis that each f_i is contractible in Q_1 so that X_2 is of the form $X_1 + H$, $H \in \mathfrak{H}(n,k,2)$ (following notation of GPC).

The g_i's induce a homomorphism $G_r \to \pi_2(Q_2)$. Let ϕ be the composition

$$G_r \to \pi_2(Q_1) \to \pi_2(X_2) \to \pi_2(H)$$

where the last homomorphism it obtained by identifying X_1 to a point in X_2.

Assertion. ϕ is an epimorphism.

Suppose the assertion is false and $\alpha \in \pi_2(H)$ is not in the image of ϕ. Then since $\pi_2(X_2) = \pi_2(X_1) + \pi_2(H)$ (by 2.6), the image of α under $\pi_2(H) \to \pi_2(X_2) \to \pi_2(X_3)$ is not in the image of $\pi_2(X_1) \to \pi_2(X_2) \to \pi_2(X_3)$. But the last composition is an isomorphism since $X_1 = M_1 \times I$, thus contradicting the existence of such an α. Hence the assertion is true.

Let $\gamma_1, \cdots, \gamma_k$ be the generators of $\pi_2(H)$ corresponding to $f_1, \cdots, f_k$. Then by 4.1 of GPC there is an automorphism β of G_r such that $\phi\beta(g_i) = \gamma_i$, $i \leqq k$, $\phi\beta(g_i) = 0$, $i > k$. By 2.1 of GPC it can be assumed that the g_i are such that $\phi(g_i) = \gamma_i$, $i \leqq k$, $\phi(g_i) = 0$, $i < k$.

Now apply 2.5 with W, V, M corresponding to

$$\chi(X_2, Q_2; g_k), \chi(X_1, Q_1; f_1, \cdots, f_k), \chi(X_1, Q_1; f_1, \cdots, f_{k-1})$$

respectively. This eliminates the critical point of f corresponding to f_k and by induction all the critical points of index two are eliminated.

Applying some of the previous considerations to $n - f$ we eliminate the critical points of f of index n, $n - 1$.

Now more generally suppose f on X_{p-1} has no critical points where $p \leqq n - 3$. Then since $H_p(X_{p+1}, X_p) = 0$, 2.1 applies to eliminate the critical points of index p. Thus we obtain by induction a function f on W with critical points only of index $n - 2$, and which satisfies the conditions a)-c) above. By 7.5 of GPC, f has no critical points at all. This proves 3.1.

COROLLARY 3.2. *Suppose W^n is compact, $n > 5$, ∂W the disjoint union of closed manifolds M_1, M_2 with each $M_i \to W$ a homotopy equivalence. Suppose also $V \subset W$ with*

$$\partial V = V_1 \cup V_2,\ V_i \subset M_i,\ V = V \times I \text{ and } \pi_1(M_i - V_i) = 1.$$

Then $i\colon \to W$ can be extended to a diffeomorphism of $M_1 \times I$ onto W.

Proof of 3.2. First i may be extended to $T \times I$ where T is a tubular neighborhood of V_1 in M_1. Then apply 3.1 to $W - V$ to get 3.2.

We now can prove 1.4. First by 3.2 with V empty applied to V of 1.4 yields that V is diffeomorphic to $V_1 \times I$. Now 3.2 applies to yield 1.4.

Section 4. The following is quit a general theorem and in fact contains 1.1 as a special case with $k = n - 1$.

THEOREM 4.1. *Suppose* $W^n \supset M^k$ *where* W *is a compact connected manifold and* M *is a closed manifold. Furthermore suppose*

(a) $\pi_1(\partial W) = \pi_1(M) = 1$

(b) $n > 5$

(c) *The inclusion of* M *into* W *is a homotopy equivalence. Then* W *is diffeomorphic to a closed cell bundle over* M, *in particular to a tubular neighborhood of* M *in* W.

We need a lemma.

LEMMA 4.2. *Suppose* B *is a compact connected n-dimensional submanifold of a compact connected manifold* V^n *with* $\partial B \cap \partial V = \emptyset$, $\pi_1(\partial B) = \pi_1(\partial V) = 1$ *and* $H_*(B) \to H_*(V)$ *induced by inclusion is bijective. Then* $Q = Cl(V - B)$ *has boundary consisting of* ∂V, ∂B *with the inclusions of* ∂V, ∂B *into* Q *homotopy equivalences.*

For the proof of 4.2 we use the following version of the Poincaré Duality Theorem, which follows from the Lefschetz Duality Theorem.

THEOREM 4.3. *Suppose* W^n *is a compact manifold* ∂W *the disjoint union of manifolds* M_1 *and* M_2 *(possibly either or both empty). Then for all* i, $H^i(W, M_1)$ *is isomorphic to* $H_{n-i}(W, M_2)$.

To prove 4.2 note $H_i(Q, \partial B) = H_i(V, B) = 0$ and $H^i(Q, \partial B) = H^i(V, B) = 0$ for all i. By 4.3 then $H_i(Q, \partial V) = 0$ for all i also. By the Whitehead theorem we get 4.2.

The proof of 4.1 then goes as follows. We can first suppose that M is disjoint from the boundary of W. Now let T be a tubular neighborhood of M which is also disjoint from ∂W. Now apply 4.2 and 3.2 to $Cl(W - T)$ with V of 3.2 empty. This yields tht $Cl(W - T)$ is diffeomorphic to $\partial T \times I$ and hence W is diffeomorphic to T. We have proved 4.1.

THEOREM 4.4. *Suppose* $2n \geqq 3m + 3$ *and a compact manifold* W^n *has the homotopy type of a closed manifold* M^m, $n > 5$, *with* $\pi_1(\partial W) = \pi_1(M) = 1$. *Then* W *is diffeomorphic to a cell-bundle over* M.

Proof. Let $f\colon M \to W$ be a homotopy equivalence. By Haefliger [1], f is homotopic to an embedding $g\colon M \to W$. Now 4.1 applies to yield 4.4.

Section 5. We continue with some consequences of 4.1. The next theorem is a strong form of the Generalized Poincaré Conjecture for $n > 5$ and was first proved in [14] except for $n = 7$. This theorem follows from 4.1 by taking M to be a point.

THEOREM 5.1. *Suppose C^n is a compact contractible manifold with $\pi_1(\partial C) = 1$ and $n > 5$. Then C is diffeomorphic to the n-disk D^n.*

For $n = 5$, if one knows in addition that ∂C is diffeomorphic to S^4, then using the theorem of Milnor $\Theta^5 = 0$ and 1.1, one obtains that C is diffeomorphic to D^5.

The following is a weak unknotting theorem in the differentiable case. Haefliger [2] has given an imbedding (differentiable) of S^3 in S^6 which does not bound an imbedded D^4. On the other hand we have:

THEOREM 5.2. *Suppose $S^k \subset S^n$ with $n - k > 2$. Then the closure of the complement of a tubular neighborhood T of S^k in S^n is diffeomorphic to $S^{n-k-1} \times D^{k+1}$.*

The proof of 5.2 is as follows (the case $n \leqq 5$ is essentially contained in Wu Wen Tsun [19]). It is well-known and easy to prove that if $X = Cl(S^n - T)$, X has the homotopy type of S^{n-k-1}. In fact T is diffeomorphic to a cell bundle over S^k and the inclusion of the boundary of a fiber S_0^{n-k-1} into X induces the equivalence. Furthermore the normal bundle of S_0^{n-k-1} in S^n is trivial because S_0^{n-k-1} bounds a disk in S^n. Now 4.1 applies to yield Theorem 5.2.

One can also prove some recent theorems of M. Hirsch [5], replacing his combinatorial arguments by application of the above theorems.

THEOREM 5.3 (Hirsch). *If $f\colon M_1^n \to M_2^n$ is a homotopy equivalence of simply connected closed manifolds such that the tangent bundle of M_1 is equivalent to the bundle over M_1 induced from the tangent bundle of M_2 by f, then $M_1 \times D^k$ and $M_2 \times D^k$ are diffeomorphic for $k > n$.*

One obtains 5.3 by imbedding M_1 in $M_2 \times D^k$ approximating the homotopy equivalence and applying 4.1. The tangential property of f is used to conclude that a tubular neighborhood of M_1 in $M_2 \times D^k$ is a product neighborhood.

THEOREM 5.4 (Hirsch). *If the homotopy sphere M^n bounds a parallelizable manifold, then $M^n \times D^3$ is diffeomorphic to $S^n \times D^3$.*

One first proves that M^n can be imbedded in S^{n+3} with trivial normal bundle by following Hirch [4] or using "handlebody theory." Then apply the argument in 5.2 to obtain the complement of a tubular neighborhood of M^n is diffeomorphic to $S^2 \times D^{n+1}$. The closure of the complement of $S^2 \times D^{n+1}$ in S^{n+3} is $S^n \times D^3$, thus proving 5.4.

Section 6. The main goal of this section is the following theorem.

THEOREM 6.1. *Let M be a simply connected closed manifold of dimension greater than five. Then on M there is a non-degenerate C^∞ function with the minimal number of critical points consistent with the homology structure.*

One actually obtains such a function with the additional property that at a critical point its value is the index.

6.2. We make more explicit the conclusion of 6.1. Suppose for each $0 \leqq i \leqq n$, $\sigma_{i1}, \cdots, \sigma_{ip(i)}, \tau_{i1}, \cdots, \tau_{iq(i)}$ is a set of generators for a corresponding direct sum decomposition of $H_i(M)$, σ_{ij} free, τ_{ij} of finite order. Then one can obtain the function of 6.1 with type numbers satisfying

$$M_i = p(i) + q(i) + q(i-1).$$

By taking the $q(i)$ minimal, the M_i becomes minimal.

In the case there is no torsion in the homology of M, 6.1 becomes.

THEOREM 6.3. *Let M be a simply connected closed manifold of dimension greater than five with no torsion in the homology of M. Then there is a non-degenerate function on M with type numbers equal the betti numbers of M.*

We start the proof of 6.1 with the following lemma.

LEMMA 6.4. *Let M^n be a simply connected compact manifold, $n > 5$, $n \geqq 2m$. Then there is an n-dimensional simply connected compact manifold X_m such that:*

a) $H_j(X_m) = 0$, $j > m$

b) *There is a "nice" function on X_m, minimal with respect to its homology structure. In other words there is a C^∞ non-degenerate function on X_m, value at a critical point equal the index, equal to $m + \frac{1}{2}$ on ∂X_m, regular in a neighborhood of ∂X_m and the k-th type number M_k is minimal in the sense of* 6.2.

c) *There is an imbedding $i \colon X_m \to M^n$ such that*

$$i(\partial X_m) \cap = M = \emptyset, \qquad i_* \colon H_j(X_m) \to H_j(X_m)$$

is bijective for $j < m$ and surjective for $j = m$.

Proof. The proof goes by induction on m, starting by taking X_1 to be an n-disk. Suppose, X_{k-1}, $i_0\colon X_{k-1}\to M$ have been constructed satisfying a)-c) with $2k\leqq n$. For convenience we identify points under i_0 so that $X_{k-1}\subset M$. We now construct X_k, $i\colon X_k\to M$ satisfying a)-c).

By the relative Hurewicz theorem the Hurewicz homomorphism

$$h\colon \pi_k(M,X_{k-1})\to H_k(M,X_{k-1})$$

is bijective.

For the structure of $H_k(M,X_{k-1})$ consider the exact sequence

$$0\to H_k(M)\to H_k(M,X_{k-1})\to H_{k-1}(X_{k-1})\xrightarrow{j} H_{k-1}(M)\to 0.$$

Let $\gamma_1,\cdots,\gamma_p$ be a set of generators of $H_k(M,X_{k-1})$ corresponding to a minimal set of generators of $H_k(M)$ together with a minimal set for $\operatorname{Ker} j$.

Represent the elements $h^{-1}(\gamma_1)\cdots h^{-1}(\gamma_p)$ by imbeddings

$$\bar{g}_i\colon (D^k,\partial D^k)\to (Cl(M-X_{k-1},\partial X_{k-1})$$

with $\bar{g}_i(D^k)$ transversal to ∂X_{k-1} along $\bar{g}_i(\partial D^k)$, for example following Wall [17], proof of Theorem 1.

In the extreme case $n=2k$, the images of $\bar{g}_i$ generically intersect each other in isolated points. These points can be removed by pushing them along arcs past the boundaries. Still following [17], the $\bar{g}_i$ can be extended to tubular neighborhoods,

$$g_i\colon (D^k,\partial D^k)\times D^{n-k}\to (Cl(M-X_{k-1}),\partial X_{k-1}).$$

Then we take X_k to be $\chi(X_{k-1};g'_1,\cdots,g'_p;k)$ where $g'_i\colon \partial D^k\times D^{n-k}\to\partial D_{k-1}$ is the restriction of g_i. It is not difficult to check that X_k has the desired properties a)-c). This proves 6.4.

To prove 6.1, let M^n be as in 6.1 with $n=2m$ or $2m+1$. Let $X_m\subset M$ as in 6.4, f the nice function on X_m and $K=Cl(M-X_m)$. Then $H_i(M,X)=0$, $i\leqq m$, so by duality $H^j(K)=0$, $j\geqq n-m$. By the Universal Coefficient Theorem this implies that $H_{n-m-1}(K)$ is torsion free. Let $Y_{n-m-1}\subset K$ be again given by 6.4 with g the nice function on Y_{n-m-1}. By 4.2 and 3.2 we can in fact assume that K and Y_{n-m-1} are the same, so $M=X_m\cup Y_{n-m-1}$. Let f_0 be the function which is f on X_m and $n-g$ on Y_{n-m-1}. By smoothing f_0 along ∂X_m we obtain a C^∞ function f'. It is not difficult using the Universal Coefficient Theorem and Poincaré Duality to show that f' may be taken as the desired function of 6.1.

The previous results of this section may be extended to manifolds with boundary.

By the previous methods one may prove the following generalization of 6.1. We leave the details to the reader.

THEOREM 6.5. *Suppose W^n is a simply connected manifold with simply-connected boundary, $n > 5$. Then there is a nice function f on W (non-degenerate, value $n + \frac{1}{2}$ on ∂W, regular in a neighborhood of ∂W, value at a critical point is the index) with type numbers minimal with respect to the homology structure of $(W, \partial W)$.*

Section 7. The goal of this section is to prove the following.

THEOREM 7.1. *Let $f\colon W_1{}^n \to W_2{}^n$ be a homotopy equivalence between two manifolds such that the tangent bundle T_1 of W_1 is equivalent to $f^{-1}T_2$. Suppose also $n > 5$, $n \geqq 2m + 1$, $H^i(W_1) = 0$, $i > m$, $\pi_1(W_1) = \pi_1(\partial W_1) = \pi_1(\partial W_2) = 1$. Then W_1 and W_2 are diffeomorphic by a diffeomorphism homotopic to f.*

Let g be a nice function on W_1 with no critical points of index greater than m, whose existence is implied by 6.5. Thei we let $X_k = g^{-1}[0, k + \frac{1}{2}]$, $k = 0, 1, \cdots, m$ with $X_m = W_1$. By 3.2 and 4.2 is is sufficient to imbed X_m in W_2 by a map homotopic to f.

Suppose inductively we have defined a map $f_{k-1}\colon X_k \to W_2$ homotopic to f with the property f_{k-1} is an imbedding, $k \geqq m$. Let X_k be written in the form $\chi(X_{k-1}; g_1, \cdots, g_p; k)$ where $g_i\colon \partial D^k \times D^{n-k} \to \partial X_{k-1}$. Using the Whitney imbedding theory we can find $f'_{k-1}\colon X_k \to W_2$ homotopic to f_{k-1}, which is an imbedding on X_{k-1} and on the images $g_i(D^k \times 0)$ in X_k as well. It remains to make f'_{k-1} an imbedding on a tubular neighborhood of each of the $g_i(D^k \times 0)$, or equivalently on each of the $g_i(D^k \times D^{n-k})$.

This can be done for a given i if and only if an element γ_i in $\pi_{k-1}(0(n-k))$ defined by f'_{k-1} in a neighborhood of $g_i(\partial D^k \times 0)$ is zero. But the original tangential assumptions on f insure $\gamma_i = 0$ in this dimension range. The arguments in proving these statements are so close to the arguments in Hirsch [3] Section 5, that we omit them. This finishes the proof of 7.1.

Section 8. We note here the following theorem.

THEOREM 8.1. *Let M^{2m+1} be a closed simply connected manifold, $m > 2$, with $H_m(M)$ torsion free. Then there is a compact manifold W^{2m+1}, uniquely determined by M and a diffeomorphism $h\colon \partial W \to \partial W$ such that M is union of two copies of W with points identified under h.*

Proof. Let $W_1^{2m+1} \subset M$ be the manifold given by 6.4. Let $W_2^{2m+1} \subset M - W$ be also given by 6.4. Then it is not difficult using homotopy theory to show that W_1, W_2 satisfy the hypotheses of 7.1. Also by previous arguments W_2 is diffeomorphic to $Cl(M - W_1)$. The uniqueness of $W_1 = W_2$ is also given by 7.1. Putting these facts together we get 8.1.

Remark. I don't believe the condition on $H_m(M)$ is really necessary here. Also in a different spirit, 8.1 is true for the cases $m = 1$, $m = 2$.

REFERENCES.

[1] A. Haefliger, "Plongement Differentiable des varietes des varietes," *Commentari Mathematicii Helvetica, to appear.*

[2] ———, "Knotted $(4k-1)$ spheres in $6k$ space," *to appear.*

[3] M. W. Hirsch, "Immersions of manifolds," *Transactions of the American Mathematical Society*, vol. 93 (1959), pp. 242-276.

[4] ———, "On manifolds which bound in Euclidean space," *to appear.*

[5] ———, *Diffeomorphisms of tubular neighborhoods* (mimeographed), Berkeley, 1961.

[6] E. T. Hu, *Homotopy theory*, New York, 1959.

[7] M. Kervaire and J. Milnor, "Groups of homotopy spheres, I and II," *to appear.*

[8] B. Mazur, "A note on some contractible 4-manifolds," *Annals of Mathematics*, vol. 73 (1961), pp. 221-228.

[9] J. Milnor, *Differentiable manifolds which are homotopy spheres* (mimeographed), Princeton, 1959.

[10] ———, *Two complexes which are homeomorphic but combinatorial distinct* (mimeographed), Princeton University (1961).

[11] A. Shapiro, "Obstructions to the imbedding of a complex in a Euclidean space, I. The first obstruction," *Annals of Mathematics*, vol. 66 (1957), pp. 256-269.

[12] S. Smale, "On gradient dynamical systems," *Annals of Mathematics*, vol. 74 (1961), pp. 199-206.

[13] ———, "Generalized Poincaré's conjecture in dimensions greater than four," *Annals of Mathematics, to appear.*

[14] ———, "Differentiable and combinatorial structures on manifolds," *Annals of Mathematics, to appear.*

[15] ———, "On the structure of five manifolds," *Annals of Mathematics, to appear.*

[16] R. Thom, "Les classes caracteristiques de Pontrjagin des varietes triangulees," *Topologia Algebraico*, Mexico, 1958, pp. 54-67.

[17] C. T. C. Wall, "Classification of $(n-1)$-connected $2n$-manifolds," *to appear.*

[18] H. Whitney, "The self-intersections of a smooth n-manifold in $2n$-space," *Annals of Mathematics*, vol. 45 (1944), pp. 220-246.

[19] Wu Wen Tsun, "On the isotopy of C^r manifolds of dimension n in Euclidean $(2n+1)$-space," *Science Record*, New Series, vol. II (1958), pp. 271-275.

ON THE REDUCTION OF INDUCED REPRESENTATIONS OF FINITE GROUPS.*[1]

By PATRICIA A. TUCKER.[2]

1. **Introduction.** If G is a finite group with a normal subgroup H, then the theory of induced modules provides an effective method of constructing a left $K(G)$-module from a left $K(H)$-module. This leads to the following question: If L is an irreducible left $K(H)$-module,[3] what are the components of the induced module $L^G = K(G) \otimes_{K(H)} L$? In this paper, the indecomposable components of L^G are determined when G is a split extension of H by G/H and K is an algebraically closed field of characteristic $p \geqq 0$.

Let G be the split extension of H by B, $G = BH$, and let T be an irreducible representation of H with representation module L. Then, $S = \{b \in B \mid T^{(b)}$ is equivalent to $T\}$ is a subgroup of B. T determines a projective representation of S with factor set α. Each left ideal of the α^{-1}-twisted group algebra, $(KS)_{\alpha^{-1}}$, (defined in § 3) determines a left $K(G)$-submodule of L^G. The main result of § 3 is a general intertwining theorem for the left ideals of $(KS)_{\alpha^{-1}}$ and the corresponding left $K(G)$-submodules of L^G. This, in particular, implies that a direct decomposition of $(KS)_{\alpha^{-1}}$ into indecomposable ideals determines a direct decomposition of L^G into indecomposable components.

If the characteristic p of K does not divide the order of G, Mackey has proven in [7] that the irreducible representations of G, an arbitrary extension of H, are of the form $(N' \otimes M)^G$ where N' and M are irreducible projective representations of a subgroup $H_0 \supseteq H$ and N' can be considered as a projective representation of H_0/H. The results obtained here for the case in which $p \nmid [G:1]$ can be used to prove Mackey's result for finite groups which are split extensions and can probably be obtained from the results in Mackey's

---

* Received October 30, 1961.

[1] This paper is based on the author's dissertation written under the direction of Dr. Charles W. Curtis and submitted to the University of Wisconsin in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research was supported in part by the Office of Naval Research and by the National Science Foundation.

[2] The author wishes to express her appreciation to Dr. Charles W. Curtis for his guidance and encouragement.

[3] All modules will be assumed to be unital and finite dimensional vector spaces over K.

paper. However, it is of interest to give an explicit method for the splitting up of T^G into irreducible components. Such a result is more satisfactory for the purpose of computing the irreducible representations of G. The intertwining theorem and the direct decomposition of T^G into indecomposable components seem to be outside the scope of Mackey's result.

The method of construction presented here can be extended to the case in which G is not a split extension of H by G/H. This result will appear later.

2. Background.

The basic definitions and notation used in this paper may be found in Curtis and Reiner [3]. The following version of Schur's Lemma (see [3, § 27]) will be used frequently in this paper.

THEOREM 1. (Schur's Lemma). *Let A be a finite dimensional algebra over an algebraically closed field K and let L and L' be irreducible left A-modules. Then,* $\mathrm{Hom}_A(L, L) = K \cdot 1_L$, *and if L and L' are not A-isomorphic, then* $\mathrm{Hom}_A(L, L') = 0$.

The following is a basic result of the theory of indecomposable modules. (See Artin, Nesbitt, and Thrall [1, p. 97].)

THEOREM 2. *Let A be a finite dimensional algebra over a field K. A left A-module L is indecomposable if and only if* $\mathrm{Hom}_A(L, L)$ *is completely primary, i. e., the residue class ring of* $\mathrm{Hom}_A(L, L)$ *modulo its radical is a division ring.*

In [6], Mackey has proved:

THEOREM 3. *Suppose K is an algebraically closed field. Let H be a normal subgroup of G and let L be an irreducible left $K(H)$-module. Then,* $\mathrm{Hom}_{K(G)}(L^G, L^G) = K \cdot 1_{L^G}$ *if and only if for every $x \notin H$, L and $x \otimes L$ are non-isomorphic left $K(H)$-modules.*

Thus, if for every $x \notin H$, L and $x \otimes L$ are non-isomorphic left $K(H)$-modules, then L^G is indecomposable.

3. The case of a split extension.

Let G be a split extension of H by B. Then, H and B are subgroups of G such that H is normal in G, $B \cap H = \{1\}$, and $G = BH$. Each $b \in B$ defines an automorphism $\bar{b}$ of H, $\bar{b} : h \to b^{-1}hb = h^b$, for every $h \in H$.

Let T be an irreducible representation of H over an algebraically closed field K of characteristic $p \geqq 0$. Let L be a representation module for T. Then L is a left $K(H)$-module.

If $x \in G$, then $T^{(x)}$, the conjugate of the representation T under x, is the representation of H with representation space L such that $T^{(x)}: h \to T^{(x)}(h) = T(x^{-1}hx)$, for every $h \in H$. $T^{(x)}$ is equivalent to T if and only if there exists a non-singular linear transformation D_x of L such that $D_x T^{(x)}(h) D_x^{-1} = T(h)$, for every $h \in H$. This will be written $D_x T^{(x)} D_x^{-1} = T$. The following properties hold for x and y in G:

i) If $D_x T^{(x)} D_x^{-1} = T$, then $D_x^{-1} T^{(x^{-1})} D_x = T$.

ii) If $D_x T^{(x)} D_x^{-1} = T$ and $D_y T^{(y)} D_y^{-1} = T$, then $D_x D_y T^{(xy)} D_y^{-1} D_x^{-1} = T$.

iii) If $D_x T^{(x)} D_x^{-1} = T$ and $C_x T^{(x)} C_x^{-1} = T$, then $D_x = \rho(x) C_x$ for some $\rho(x) \in K^* = \{k \in K \mid k \neq 0\}$.

The proofs of i) and ii) follow immediately from the definitions and will be omitted. In iii), it follows from i) that $C_x^{-1} T^{(x^{-1})} C_x = T$ and then by ii) that

$$T = D_x C_x^{-1} T^{(xx^{-1})} C_x D_x^{-1} = D_x C_x^{-1} T C_x D_x^{-1}.$$

But T is irreducible, thus by Schur's Lemma (Theorem 1)

$$D_x C_x^{-1} = \rho(x) 1_L, \text{ for some } \rho(x) \in K^*.$$

Thus, $D_x = \rho(x) C_x$ and iii) is proved.

Let $S = \{b \in B \mid T^{(b)} \text{ is equivalent to } T\}$. By i) and ii), S is a subgroup of B. It follows from Theorem 3 using i) and ii), that if $S = \{1\}$, then T^G is indecomposable. Suppose $S \neq \{1\}$. For every $b \in S$, select a linear transformation D_b of L such that $D_b T^{(b)} D_b^{-1} = T$ and $D_1 = 1_L$. Then, for every $l \in L$, $h \in H$, $D_b h^b D_b^{-1} l = hl$.

Consider the map σ of S into $\mathrm{Hom}_K(L, L)$ defined by $\sigma: b \to D_b$. Then, $\sigma(1) = 1_L$. Let $a, b \in S$. By ii), $D_a D_b T^{(ab)} D_b^{-1} D_a^{-1} = T$. But $D_{ab} T^{(ab)} D_{ab}^{-1} = T$. Thus, by iii),

$$D_a D_b = \alpha(a, b) D_{ab}$$

for some $\alpha(a, b) \in K^*$ Then,

$$\sigma(a)\sigma(b) = \alpha(a, b)\sigma(ab).$$

Therefore, σ is a projective representation of S.

It can be shown using iii), that if σ' is any projective representation of S in $\mathrm{Hom}_K(L, L)$ such that $\sigma'(b) T^{(b)} \sigma'(b)^{-1} = T$, then σ' is equivalent to σ. Then, the factor set α' of σ' is equivalent to α. Conversely, let α' be a factor set of S which is equivalent to α. Then, there exists a function $\rho\colon S \to K^*$ such that for every $a, b \in S$,

$$\alpha'(a,b) = \alpha(a,b)\rho(ab)\rho(a)^{-1}\rho(b)^{-1}.$$

Define σ', a map of S into $\mathrm{Hom}_K(L,L)$, by

$$\sigma'\colon b \to C_b = \rho(b)^{-1}D_b = \rho(b)^{-1}\sigma(b).$$

Then, σ' is a projective representation of S with factor set α' and σ' is equivalent to σ. Thus T uniquely determines an equivalence class of factor sets of S.

Let α be a factor set of S. The α-twisted group algebra of S, denoted by $(KS)_\alpha$, is the crossed product of K and S with factor set α. $(KS)_\alpha$ consists of all formal sums $\sum_{b \in S} \xi(b)b$, where the $\xi(b) \in K$. By definition, the elements of $(KS)_\alpha$ will satisfy the following:

1) $\sum_{b \in S} \xi(b)b = \sum_{b \in S} \eta(b)b$ if and only if $\xi(b) = \eta(b)$,

   for every $b \in S$.

2) $k(\sum_{b \in S} \xi(b)b) + k'(\sum_{b \in S} \eta(b)b) = \sum_{b \in S} (k\xi(b) + k'\eta(b))b$,

   for all $k, k' \in K$.

3) $(\sum_{b' \in S} \xi(b')b') \cdot (\sum_{b \in S} \eta(b)b) = \sum_{b', b \in S} \xi(b')\eta(b)\alpha(b',b)b'b$

In particular, $b' \cdot b = \alpha(b',b)b'b$, for b', $b \in S$.

The elements of S form a basis for $(KS)_\alpha$ as a vector space over K. $(KS)_\alpha$ considered as a left $(KS)_\alpha$-module affords a projective representation of S with factor set α. $(KS)_\alpha$ and $(KS)_{\alpha'}$ are equivalent, i.e., they afford equivalent projective representations of S, if and only if α and α' are equivalent factor sets of S. There is a one-to-one correspondence between the projective representations of S with factor set α and the left $(KS)_\alpha$-modules.

If the characteristic p of K does not divide $[S\colon 1]$, then any projective representation of S is completely reducible. (See Jacobson [4, p. 81].) Then, $(KS)_\alpha$ is semi-simple and every irreducible $(KS)_\alpha$-module is $(KS)_\alpha$-isomorphic to some minimal left ideal of $(KS)_\alpha$.

Define the factor set α^{-1} of S by

$$\alpha^{-1}(b,b') = \alpha(b,b')^{-1},$$

for every $b, b' \in S$. If α' is equivalent to α, it follows from the definition of equivalence that $(\alpha')^{-1}$ is equivalent to α^{-1}. Thus, T determines up to equivalence the α^{-1}-twisted group algebra $(KS)_{\alpha^{-1}}$ of S where α is the factor set of the projective representation σ such that $\sigma\colon b \to D_b$. $(KS)_{\alpha^{-1}}$ will be used in the construction which follows. The elements of $(KS)_{\alpha^{-1}}$ will be denoted by s, where

$$s = \sum_{b \in S} \xi_s(b) b.$$

Let a_γ, $\gamma = 1, \cdots, r = [B\colon S]$ and $a_1 = 1$, be a complete set of left coset representatives of S in B. Thus $B = S + a_2 S + \cdots + a_r S$. It is easily checked that $T^{(a_\gamma)}$ and $T^{(a_\delta)}$ are inequivalent for $\gamma \neq \delta$. The left $K(G)$-module L^G for T^G has dimension $[B\colon 1](L\colon K)$ and is generated as a vector space over K by the elements

$$\{a_\gamma b \otimes l \mid 1 \leqq \gamma \leqq r,\ b \in S, l \in L\}.$$

For every $\gamma = 1, \cdots, r$, $s \in (KS)_{\alpha^{-1}}$, and $l \in L$, define an element $c(a_\gamma, s, l) \in L^G$ by

$$c(a_\gamma, s, l) = \sum_{b \in S} \xi_s(b) a_\gamma b \otimes D_b^{-1} l$$

The $c(a_\gamma, s, l)$'s satisfy the following properties:

PROPERTY 1. $c(a_\gamma, s + s', l) = c(a_\gamma, s, l) + c(a_\gamma, s', l)$.

PROPERTY 2. $c(a_\gamma, s, l + l') = c(a_\gamma, s, l) + c(a_\gamma, s, l')$.

PROPERTY 3. *If $k \in K$, then*
$$k(c(a_\gamma, s, l)) = c(a_\gamma, ks, l) = c(a_\gamma, s, kl).$$

PROPERTY 4. *If $h \in H$, then $h(c(a_\gamma, s, l)) = c(a_\gamma, s, h^{a_\gamma} l)$.*

PROPERTY 5. *If $b \in S$, then $b(c(a_\gamma, s, l)) = c(a_\delta, b' \cdot s, D_{b'} l)$, where*
$$b a_\gamma = a_\delta b',\ b' \cdot s = \sum_{b \in S} \xi_s(b) \alpha(b', b)^{-1} b' b.^{4}$$

PROPERTY 6. *If a_β is a coset representative of S in B, then $a_\beta(c(a_\gamma, s, l)) = c(a_\delta, b' \cdot s, D_{b'} l)$, where $a_\beta a_\gamma = a_\delta b'$, and $b' \cdot s = \sum_{b \in S} \xi_s(b) \alpha(b', b)^{-1} b' b$.*

The proofs of Properties 1, 2, and 3 will be omitted. The proof of Property 6 is similar to the proof of 5 and thus will be omitted.

Proof of Property 4. $h(c(a_\gamma, s, l)) = h(\sum_{b \in S} \xi_s(b) a_\gamma b \otimes D_b^{-1} l)$

$$= \sum_{b \in S} \xi_s(b) a_\gamma b h^{a_\gamma b} \otimes D_b^{-1} l = \sum_{b \in S} \xi_s(b) a_\gamma b \otimes h^{a_\gamma b} D_b^{-1} l$$

$$= \sum_{b \in S} \xi_s(b) a_\gamma b \otimes D_b^{-1} h^{a_\gamma} l = c(a_\gamma, s, h^{a_\gamma} l).$$

---

[4] $b' \cdot s$ denotes the product of b' and s in $(KS)_{\alpha^{-1}}$.

Proof of Property 5. $b(c(a_\gamma, s, l)) = b(\sum_{b \in S} \xi_s(b) a_\gamma b \otimes D_b^{-1} l)$

$$= \sum_{b \in S} \xi_s(b) a_\delta b' b \otimes D_b^{-1} l = \sum_{b \in S} \xi_s(b) a_\delta b' b \otimes D_b^{-1} D_{b'}^{-1} D_{b'} l$$

$$= \sum_{b \in S} \xi_s(b) a_\delta b' b \otimes (D_{b'} D_b)^{-1} D_{b'} l$$

$$= \sum_{b \in S} \xi_s(b) a_\delta b' b \otimes (\alpha(b', b) D_{b'b})^{-1} D_{b'} l$$

$$= \sum_{b \in S} \xi_s(b) \alpha(b', b)^{-1} a_\delta b' b \otimes D_{b'b}^{-1} D_{b'} l$$

$$= c(a_\delta, b' \cdot s, D_{b'} l).$$

Let I be a left ideal of $(KS)_{\alpha^{-1}}$. Define $C(I)$ to be the K-subspace of L^G generated by the elements

$$\{c(a_\gamma, s, l) \mid 1 \leqq \gamma \leqq r, s \in (KS)_{\alpha^{-1}}, l \in L\}.$$

Let $\{s_k = \sum_{b \in S} \zeta_k(b) b \mid 1 \leqq k \leqq (I:K)\}$ be a basis for I and $\{l_\mu \mid 1 \leqq \mu \leqq (L:K)\}$ be a basis for L. By the linearity of the $c(a_\gamma, s, l)$'s in s and l (Properties 1, 2, and 3), $C(I)$ is generated by the elements

$$\{c(a_\gamma, s_k, l_\mu) \mid 1 \leqq \gamma \leqq r, 1 \leqq k \leqq (I:K), 1 \leqq \mu \leqq (L:K)\}. \tag{3.1}$$

In order to prove that the elements (3.1) form a basis for $C(I)$, consider the linear transformation M of L^G such that for every b, γ, and μ,

$$M(a_\gamma b \otimes l_\mu) = a_\gamma b \otimes D_b^{-1} l_\mu.$$

It can be shown that M is a non-singular linear transformation of L^G. Thus, the elements

$$\{a_\gamma b \otimes D_b^{-1} l_\mu \mid 1 \leqq \gamma \leqq r, b \in S, 1 \leqq \mu \leqq (L:K)\}$$

form a basis for L^G.

Suppose

$$\sum_{\gamma,k,\mu} \beta(\gamma, k, \mu) c(a_\gamma, s_k, l_\mu) = 0.$$

Then,

$$\sum_{\gamma,k,\mu} \beta(\gamma, k, \mu) (\sum_{b \in S} \zeta_k(b) a_\gamma b \otimes D_b^{-1} l_\mu)$$

$$= \sum_{\gamma,k,\mu,b} \beta(\gamma, k, \mu) \zeta_k(b) (a_\gamma b \otimes D_b^{-1} l_\mu) = 0.$$

Since the $a_\gamma b \otimes D_b^{-1} l_\mu$ are linearly independent,

$$\sum_k \beta(\gamma, k, \mu) \zeta_k(b) = 0$$

for every γ, b, and μ.

But then, for every γ and μ,

$$\sum_k \beta(\gamma,k,\mu)s_k = \sum_{k,b} \beta(\gamma,k,\mu)\zeta_k(b)b = 0.$$

But the s_k are linearly independent since they form a basis for I. Thus,

$$\beta(\gamma,k,\mu) = 0, \text{ for every } \gamma,\ k, \text{ and } \mu.$$

Therefore, the elements (3.1) are linearly independent and consequently form a basis for $C(I)$. From this it follows that $(C(I):K) = [B:S](I:K)(L:K)$.

If $S \neq B$, then $C(I)$ is $K(G)$-isomorphic to a left $K(G)$-module which is induced from a left $K(SH)$-module, where SH is a proper subgroup of G. In order to show this, let $S \neq B$. Define I' to be the K-subspace of L^G generated by the elements

$$\{c(1,s,l) \mid s \in I, l \in L\}.$$

Let $SH = \{bh \mid b \in S, h \in H\}$. Since H is normal in G, SH is a subgroup of G. Since $B \cap H = \{1\}$,

$$G = BH = (a_1S + \cdots + a_rS)H = a_1SH + \cdots + a_rSH.$$

Thus, $\{a_\gamma \mid 1 \leqq \gamma \leqq r\}$ is a complete set of left coset representatives of SH in G. Let $b \in S$, $h \in H$, and $c(1,s,l) \in I'$. Then,

$$bh(c(1,s,l)) = b(c(1,s,hl)) = c(1,b\cdot s, D_bhl).$$

Therefore, $bh(c(1,s,l)) \in I'$. Thus, I' is a $K(SH)$-submodule of L^G. The elements

$$\{c(1,s_k,l_\mu) \mid 1 \leqq k \leqq (I:K), 1 \leqq \mu \leqq (L:K)\}$$

form a basis for I'. Thus, $(I')^G = K(G) \otimes_{K(SH)} I'$ has as a basis

$$\{a_\gamma \otimes c(1,s_k,l_\mu) \mid 1 \leqq \gamma \leqq r, 1 \leqq k \leqq (I:K), 1 \leqq \mu \leqq (L:K)\}.$$

Let ϕ be the linear transformation of $(I')^G$ into $C(I)$ such that for every γ, k, and μ

$$\phi\colon a_\gamma \otimes c(1,s_k,l_\mu) \to c(a_\gamma,s_k,l_\mu).$$

It is easily checked that ϕ is a $K(G)$-isomorphism of $(I')^G$ onto $C(I)$. Thus, $(I')^G$ and $C(I)$ are isomorphic as left $K(G)$-modules.

THEOREM 4. *Let* $I_1,\cdots,I_t$ *be left ideals of* $(KS)_{\alpha^{-1}}$ *such that* $(KS)_{\alpha^{-1}} = I_1 \oplus \cdots \oplus I_t$. Then

$$L^G = C(I_1) \oplus \cdots \oplus C(I_t).$$

Proof. By the definition of the $C(I_i)$,

$$C(I_1)+\cdots+C(I_t)\subseteq L^G.$$

It is sufficient to show that $a_\gamma b\otimes l\in C(I_1)+\cdots+C(I_t)$, for every a_γ, b, and l in order to prove equality. Let $s=b$. Since

$$s\in(KS)_{\alpha^{-1}},\ s=s_1+\cdots+s_t,$$

where $s_i\in I_i$. Thus, $c(a_\gamma,s_i,D_bl)\in C(I_i)$, for every i, and

$$\sum_{i=1}^{t}c(a_\gamma,s_i,D_bl)=c(a_\gamma,s,D_bl)=a_\gamma b\otimes l\in C(I_1)+\cdots+C(I_t).$$

Therefore, $L^G=C(I_1)+\cdots+C(I_t)$.

$$L^G=C(I_1)+\cdots+C(I_t)\text{ implies }(L^G\colon K)\leqq\sum_{i=1}^{t}(C(I_i):K).$$

The sum is direct if and only if $(L^G\colon K)=\sum_{i=1}^{t}(C(I_i):K)$. It has been shown that $(C(I_i):K)=[B:S](I_i:K)(L:K)$. Thus,

$$\begin{aligned}\sum_{i=1}^{t}(C(I_i):K)&=[B:S](L:K)\sum_{i=1}^{t}(I_i:K)\\&=[B:S](L:K)[S:1]=[B:1](L:K).\end{aligned}$$

But, $[B:1](L:K)=(L^G:K)$. Therefore, $(L^G:K)=\sum_{i=1}^{t}(C(I_i):K)$ and $L^G=C(I_1)\oplus\cdots\oplus C(I_t)$.

THEOREM 5. *Let I_1 and I_2 be left ideals of $(KS)_{\alpha^{-1}}$. Then there exists a K-isomorphism θ of $\mathrm{Hom}_{(KS)_{\alpha}^{-1}}(I_1,I_2)$ onto $\mathrm{Hom}_{K(G)}(C(I_1),C(I_2))$ such that if $P\in\mathrm{Hom}_{(KS)_{\alpha}^{-1}}(I_1,I_2)$, then $\theta P(c(a_\gamma,s,l))=c(a_\gamma,Ps,l)$, for every $c(a_\gamma,s,l)\in C(I_1)$.*

Proof. Let $\{s_{k1}\mid 1\leqq k\leqq(I_1:K)\}$ and $\{s_{k2}\mid 1\leqq k\leqq(I_2:K)\}$ be bases for I_1 and I_2, respectively, over K. Let $\{l_\mu\colon 1\leqq\mu\leqq(L:K)\}$ be a basis for L over K.

Let $P\in\mathrm{Hom}_{(KS)_{\alpha}^{-1}}(I_1,I_2)$. Define $\theta P=P'$ to be the K-homomorphism of $C(I_1)$ into $C(I_2)$ such that for every γ, k, and μ,

$$P'(c(a_\gamma,s_{k1},l_\mu))=c(a_\gamma,Ps_{k1},l_\mu).$$

By Property 4, P' is a $K(H)$-homomorphism of $C(I_1)$ into $C(I_2)$. Let $b\in S$. Then

$$\begin{aligned} b(P'c(a_\gamma, s_{k1}, l_\mu)) &= bc(a_\gamma, Ps_{k1}, l_\mu) \\ &= c(a_{\gamma(b)}, b(\gamma)\cdot Ps_{k1}, D_{b(\gamma)}l_\mu), \text{ where } ba_\gamma = a_{\gamma(b)}b(\gamma), \\ &= c(a_{\gamma(b)}, P(b(\gamma)\cdot s_{k1}), D_{b(\gamma)}l_\mu) = P'(c(a_{\gamma(b)}, b(\gamma)\cdot s_{k1}, D_{b(\gamma)}l_\mu)) \\ &= P'(bc(a_\gamma, s_{k1}, l_\mu)). \end{aligned}$$

Therefore, P' is a $K(S)$-homomorphism. Similarly, it can be shown that $a_\beta(P'c(a_\gamma, s_{k1}, l_\mu)) = P'(a_\beta c(a_\gamma, s_{k1}, l_\mu))$. Therefore, P' is a $K(G)$- homomorphism of $C(I_1)$ into $C(I_2)$. Thus, $\theta\colon P \rightarrow P'$ is a map of $\mathrm{Hom}_{(KS)_{a^{-1}}}(I_1, I_2)$ into $\mathrm{Hom}_{K(G)}(C(I_1), C(I_2))$.

Consider $(P+Q)'$, where $P, Q \in \mathrm{Hom}_{(KS)_{a^{-1}}}(I_1, I_2)$.

$$\begin{aligned} (P+Q)'(c(a_\gamma, s_{k1}, l_\mu)) &= c(a_\gamma, (P+Q)s_{k1}, l_\mu) \\ &= c(a_\gamma, Ps_{k1}+Qs_{k1}, l_\mu) = c(a_\gamma, Ps_{k1}, l_\mu) + c(a_\gamma, Qs_{k1}, l_\mu) \\ &= P'(c(a_\gamma, s_{k1}, l_\mu)) + Q'(c(a_\gamma, s_{k1}, l_\mu)). \end{aligned}$$

Thus, $(P+Q)' = P' + Q'$. Let $k \in K$. Then,

$$\begin{aligned} (kP)'(c(a_\gamma, s_{k1}, l_\mu)) = c(a_\gamma, kPs_{k1}, l_\mu) &= kc(a_\gamma, Ps_{k1}, l_\mu) \\ &= kP'(c(a_\gamma, s_{k1}, l_\mu)). \end{aligned}$$

Thus, $(kP)' = kP'$. Therefore, θ is a K-homomorphism of $\mathrm{Hom}_{(KS)_{a^{-1}}}(I_1, I_2)$ into $\mathrm{Hom}_{K(G)}(C(I_1), C(I_2))$.

If $I_1 = I_2 = I$, then $\mathrm{Hom}_{(KS)_{a^{-1}}}(I, I)$ and $\mathrm{Hom}_{K(G)}(C(I), C(I))$ are algebras over K. Let $P, Q \in \mathrm{Hom}_{(KS)_{a^{-1}}}(I, I)$. Then, consider $(PQ)'$.

$$\begin{aligned} (PQ)'(c(a_\gamma, s_{k1}, l_\mu)) &= c(a_\gamma, (PQ)s_{k1}, l_\mu) = c(a_\gamma, P(Qs_{k1}), l_\mu) \\ &= P'(c(a_\gamma, Qs_{k1}, l_\mu)) = P'Q'(c(a_\gamma, s_{k1}, l_\mu)). \end{aligned}$$

Thus, $(PQ)' = P'Q'$. Therefore, θ is an algebra homomorphism of $\mathrm{Hom}_{(KS)_{a^{-1}}}(I, I)$ into $\mathrm{Hom}_{K(G)}(C(I), C(I))$.

In order to prove that θ is onto and one-to-one, let

$$N \in \mathrm{Hom}_{K(G)}(C(I_1), C(I_2))$$

and let $N(c(a_\gamma, s_{k1}, l_\mu)) = \sum\limits_{\delta, j, \nu} n(a_\delta, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\mu)c(a_\delta, s_{j2}, l_\nu)$. Let $h^{a_\gamma}l_\mu = \sum\limits_{\nu=1}^{(L:K)} t_{\nu\mu}(h^{a_\gamma})l_\nu$.

If R is a linear transformation of L such that $Rh^{a_\gamma}l_\mu = h^{a_\delta}Rl_\mu$, for every $h \in H$, and $\mu = 1, \cdots, m$, then, since T^{a_γ} is irreducible and T^{a_γ} and T^{a_δ} are inequivalent for $\gamma \neq \delta$, Schur's Lemma implies

$$\begin{aligned} R &= r1_L, \quad \text{for some } r \in K, \text{ if } \gamma = \delta \\ R &= 0, \qquad\qquad\qquad\quad\;\; \text{if } \gamma \neq \delta. \end{aligned}$$

Let $Rl_\mu = \sum_{\nu=1}^{(L:K)} r_{\nu\mu} l_\nu$. Then, $Rh^{a_\gamma}l_\mu = h^{a_\delta}Rl_\mu$ is equivalent to

$$\begin{aligned} R\big(\sum_\nu t_{\nu\mu}(h^{a_\gamma})l_\nu\big) &= \sum_\nu t_{\nu\mu}(h^{a_\gamma})\big(\sum_\eta r_{\eta\nu}l_\eta\big) \\ &= h^{a_\delta}\big(\sum_\nu r_{\nu\mu}l_\nu\big) = \sum_\nu r_{\nu\mu}\big(\sum_\eta t_{\eta\nu}(h^{a_\delta})l_\eta\big). \end{aligned}$$

Thus Schur's Lemma implies

(3.2) If $\sum_{\nu=1}^{(L:K)} t_{\nu\mu}(h^{a_\gamma})r_{\eta\nu} = \sum_{\nu=1}^{(L:K)} r_{\nu\mu}t_{\eta\nu}(h^{a_\delta})$, for every η, μ, and h such that $1 \leqq \eta, \mu \leqq m$, and $h \in H$, then

$$\begin{aligned} \text{if } \gamma = \delta,\ r_{\eta\nu} &= 0 \quad \text{for } \eta \neq \nu \\ r_{\eta\eta} &= r_{\nu\nu} \quad \text{for every } \eta, \nu \\ \text{and if } \gamma \neq \delta,\ r_{\eta\nu} &= 0 \quad \text{for every } \eta, \nu. \end{aligned}$$

If $h \in H$, then for every γ, k, and μ,

$$h(Nc(a_\gamma, s_{k1}, l_\mu)) = N(hc(a_\gamma, s_{k1}, l_\mu)).$$

But

$$\begin{aligned} h(Nc(a_\gamma, s_{k1}, l_\mu)) &= h\big(\sum_{\delta,j,\nu} n(a_\delta, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\mu)c(a_\delta, s_{j2}, l_\nu)\big) \\ &= \sum_{\delta,j,\nu} n(a_\delta, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\mu)c(a_\delta, s_{j2}, h^{a_\delta}l_\nu) \\ &= \sum_{\delta,j,\nu} n(a_\delta, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\mu)\big(\sum_\eta t_{\eta\nu}(h^{a_\delta})c(a_\delta, s_{j2}, l_\eta)\big) \end{aligned}$$

and

$$\begin{aligned} N(hc(a_\gamma, s_{k1}, l_\mu)) &= N(c(a_\gamma, s_{k1}, h^{a_\gamma}l_\mu)) = N\big(\sum_\nu t_{\nu\mu}(h^{a_\gamma})c(a_\gamma, s_{k1}, l_\nu)\big) \\ &= \sum_\nu t_{\nu\mu}(h^{a_\gamma})\big(\sum_{\delta,j,\eta} n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu)c(a_\delta, s_{j2}, l_\eta)\big). \end{aligned}$$

Thus, for every γ, δ, j, k, μ, η, and for every $h \in H$,

$$\sum_\nu n(a_\delta, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\mu)t_{\eta\nu}(h^{a_\delta}) = \sum_\nu t_{\nu\mu}(h^{a_\gamma})n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu).$$

Let $r(\delta, \gamma, j, k)_{\eta\nu} = n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu)$. Then, for any choice of δ, γ, j, and k, $\{r(\delta, \gamma, j, k)_{\eta\nu} \mid 1 \leqq \eta, \nu \leqq m\}$ satisfies the conditions of (3.2). Thus, the following relationships hold.

(3.3) $n(a_\gamma, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu) = 0$ for every γ, j, k, η and ν such that $\eta \neq \nu$.

(3.4) $n(a_\gamma, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\eta) = n(a_\gamma, s_{j2}, l_\nu, a_\gamma, s_{k1}, l_\nu)$,
for every γ, j, k, η, and ν.

(3.5) $n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu) = 0$,
for every δ, γ, j, k, η, and ν such that $\delta \neq \gamma$.

Let $\gamma \neq 1$. Then, for every k and ν,

$$a_\gamma(Nc(1, s_{k1}, l_\nu)) = N(a_\gamma c(1, s_{k1}, l_\nu)) = Nc(a_\gamma, s_{k1}, l_\nu). \tag{3.6}$$

This implies that

$$\begin{aligned} a_\gamma\Big(\sum_{\delta, j, \eta} n(a_\delta, s_{j2}, l_\eta, 1, s_{k1}, l_\nu)\, c(a_\delta, s_{j2}, l_\eta)\Big) \\ = \sum_{\delta, j, \eta} n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu)\, c(a_\delta, s_{j2}, l_\eta). \end{aligned}$$

Suppose $a_\gamma a_\delta = a_\gamma b'$. Then $a_\delta = b'$ which is a contradiction unless $a_\delta = b' = 1$. Thus comparing coefficients of $c(a_\gamma, s_{j2}, l_\eta)$ in (3.6), the following relationship is obtained.

$$n(1, s_{j2}, l_\eta, 1, s_{k1}, l_\nu) = n(a_\gamma, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\nu), \tag{3.7}$$

for every γ, j, k, η, and ν.

If $b \in S$, then for every k and μ,

$$b(Nc(1, s_{k1}, l_\mu)) = N(bc(1, s_{k1}, l_\mu)) = N(c(1, b \cdot s_{k1}, D_b l_\mu)). \tag{3.8}$$

Let $b \cdot s_{ki} = \sum_{j=1}^{(I_i:K)} f^i{}_{jk}(b) s_{ji}$, $i = 1, 2$, and

$$D_b l_\mu = \sum_{\nu=1}^{(L:K)} d_{\nu\mu}(b) l_\nu.$$

Then,

$$\begin{aligned} b(Nc(1, s_{k1}, l_\mu)) &= b\Big(\sum_{\gamma, j, \nu} n(a_\gamma, s_{j2}, l_\nu, 1, s_{k1}, l_\mu)\, c(a_\gamma, s_{j2}, l_\nu)\Big) \\ &= \sum_{\gamma, j, \nu} n(a_\gamma, s_{j2}, l_\nu, 1, s_{k1}, l_\mu)\, c(a_{\gamma(b)}, b(\gamma) \cdot s_{j2}, D_{b(\gamma)} l_\nu), \end{aligned}$$

where $ba_\gamma = a_{\gamma(b)} b(\gamma)$,

$$= \sum_{\gamma, j, \nu} n(a_\gamma, s_{j2}, l_\nu, 1, s_{k1}, l_\mu) \Big(\sum_i f^2{}_{ij}(b(\gamma)) \Big) \Big(\sum_\eta d_{\eta\nu}(b(\gamma))\, c(a_{\gamma(b)}, s_{i2}, l_\eta) \Big) \Big)$$

and

$$\begin{aligned} N(c(1, b \cdot s_{k1}, D_b l_\mu)) &= N\Big(\sum_j f^1{}_{jk}(b) \Big(\sum_\nu d_{\nu\mu}(b)\, c(1, s_{j1}, l_\nu) \Big) \Big) \\ &= \sum_j f^1{}_{jk}(b) \Big(\sum_\nu d_{\nu\mu}(b) \Big(\sum_{\delta, i, \eta} n(a_\delta, s_{i2}, l_\eta, 1, s_{j1}, l_\nu)\, c(a_\delta, s_{i2}, l_\eta) \Big) \Big). \end{aligned}$$

Suppose $ba_\gamma = b(\gamma)$. Then $a_\gamma = 1$ and $b = b(\gamma)$. Thus, by comparing coefficients of $c(1, s_{i2}, l_\eta)$ in (3.8), the relationship

$$\begin{aligned} \sum_{j, \nu} n(1, s_{j2}, l_\nu, 1, s_{k1}, l_\mu) f^2{}_{ij}(b)\, d_{\eta\nu}(b) \\ = \sum_j f^1{}_{jk}(b) \Big(\sum_\nu d_{\nu\mu}(b)\, n(1, s_{i2}, l_\eta, 1, s_{j1}, l_\nu) \Big) \end{aligned} \tag{3.9}$$

is obtained for every k, i, μ, η, and b where $b \in S$. But, by (3.3), $n(1, s_{j2}, l_\nu, 1, s_{k1}, l_\mu) = 0$ for $\nu \neq \mu$, thus (3.9) reduces to

$$\sum_j n(1, s_{j2}, l_\mu, 1, s_{k1}, l_\mu) f^2{}_{ij}(b)\, d_{\eta\mu}(b) \\ = \sum_j f^1{}_{jk}(b)\, d_{\eta\mu}(b)\, n(1, s_{i2}, l_\eta, 1, s_{j1}, l_\eta),$$

for every k, i, μ, η, and b where $b \in S$.

Since D_b is non-singular, for a given μ, there exists $\eta(\mu)$ such that $d_{\eta(\mu)\mu}(b) \neq 0$. Thus,

$$\sum_j n(1, s_{j2}, l_\mu, 1, s_{k1}, l_\mu) f^2{}_{ij}(b) \\ = \sum_j f^1{}_{jk}(b)\, n(1, s_{i2}, l_{\eta(\mu)}, 1, s_{j1}, l_{\eta(\mu)}),$$

for every k, i, μ, and b where $b \in S$. But by (3.4), the $n(1, s_{i2}, l_\nu, 1, s_{j1}, l_\nu)$'s are equal for every ν, thus

$$(3.10) \qquad \sum_j n(1, s_{j2}, l_\mu, 1, s_{k1}, l_\mu) f^2{}_{ij}(b) = \sum_j f^1{}_{jk}(b)\, n(1, s_{i2}, l_\eta, 1, s_{j1}, l_\eta),$$

for every k, i, μ, η, and b where $b \in S$.

Define $p(s_{j2}, s_{k1}) = n(1, s_{j2}, l_1, 1, s_{k1}, l_1)$, for every j and k. By (3.7) and (3.4), $p(s_{j2}, s_{k1}) = n(a_\gamma, s_{j2}, l_\mu, a_\gamma, s_{k1}, l_\mu)$, for every γ and μ. By (3.10),

$$(3.11) \qquad \sum_j p(s_{j2}, s_{k1}) f^2{}_{ij}(b) = \sum_j f^1{}_{jk}(b)\, p(s_{i2}, s_{j1}), \text{ for every } k \text{ and } i.$$

Let P be the K-homomorphism of I_1 into I_2 defined by

$$P(s_{k1}) = \sum_{j=1}^{(I_2:K)} p(s_{j2}, s_{k1}) s_{j2}, \text{ for every } k.$$

Then,

$$b \cdot (P s_{k1}) = \sum_j p(s_{j2}, s_{k1})\, b \cdot s_{j2} = \sum_j p(s_{j2}, s_{k1}) \big(\sum_i f^2{}_{ij}(b) s_{i2} \big)$$

and

$$P(b \cdot s_{k1}) = P\big(\sum_j f^1{}_{jk}(b) s_{j1} \big) = \sum_j f^1{}_{jk}(b) \big(\sum_i p(s_{i2}, s_{j1}) s_{i2} \big).$$

Therefore, by (3.11), $b \cdot (P s_{k1}) = P(b \cdot s_{k1})$ and P is a $(KS)_{\alpha^{-1}}$-homomorphism.

It remains to be shown that $P' = N$.

$$\begin{aligned} N(c(a_\gamma, s_{k1}, l_\mu)) &= \sum_{\delta, j, \eta} n(a_\delta, s_{j2}, l_\eta, a_\gamma, s_{k1}, l_\mu)\, c(a_\delta, s_{j2}, l_\eta) \\ &= \sum_j n(a_\gamma, s_{j2}, l_\mu, a_\gamma, s_{k1}, l_\mu)\, c(a_\gamma, s_{j2}, l_\mu), \text{ by (3.3)-(3.5)}, \\ &= \sum_j p(s_{j2}, s_{k1})\; c(a_\gamma, s_{j2}, l_\mu) = c(a_\gamma, \sum_j p(s_{j2}, s_{k1}) s_{j2}, l_\mu) \\ &= c(a_\gamma, P s_{k1}, l_\mu) = P'(c(a_\gamma, s_{k1}, l_\mu)). \end{aligned}$$

Therefore, $P' = N$ and θ is onto.

Suppose $N = Q'$, for some $Q \in \mathrm{Hom}_{(KS)_{\alpha^{-1}}}(I_1, I_2)$, and let

$$Qs_{k1} = \sum_{j=1}^{(I_2:K)} q(s_{j2}, s_{k1}) s_{j2}.$$

Then,

$$N(c(a_\gamma, s_{k1}, l_\mu)) = Q'(c(a_\gamma, s_{k1}, l_\mu)) = c(a_\gamma, Qs_{k1}, l_\mu)$$
$$= c(a_\gamma, \sum_j q(s_{j2}, s_{k1}) s_{j2}, l_\mu) = \sum_j q(s_{j2}, s_{k1}) c(a_\gamma, s_{j2}, l_\mu).$$

But,

$$N(c(a_\gamma, s_{k1}, l_\mu)) = \sum_j p(s_{j2}, s_{k1}) c(a_\gamma, s_{j2}, l_\mu).$$

Thus, $p(s_{j2}, s_{k1}) = q(s_{j2}, s_{k1})$, for every j and k, and $Q = P$. Therefore, θ is one-to-one. This completes the proof of Theorem 5.

COROLLARY 1. *I is indecomposable if and only if $C(I)$ is indecomposable.*

Proof. By Theorem 2 and by Theorem 5 since θ is an algebra homomorphism when $I_1 = I_2 = I$, the following statements are equivalent: (i) I is indecomposable; (ii) $\mathrm{Hom}_{(KS)_{\alpha^{-1}}}(I, I)$ is completely primary; (iii) $\mathrm{Hom}_{K(G)}(C(I), C(I))$ is completely primary; (iv) $C(I)$ is indecomposable.

COROLLARY 2. *L^G is indecomposable if and only if $(KS)_{\alpha^{-1}}$ is indecomposable.*

COROLLARY 3. *If the characteristic p of K does not divide $[G:1]$, then I is irreducible if and only if $C(I)$ is irreducible.*

Proof. Since $p \nmid [G:1]$, I and $C(I)$ are completely reducible. Thus, Corollary 3 follows from Corollary 1.

COROLLARY 4. *Suppose $p \nmid [G:1]$, and let I_1 and I_2 be irreducible. Then, I_1 and I_2 are $(KS)_{\alpha^{-1}}$-isomorphic if and only if $C(I_1)$ and $C(I_2)$ are $K(G)$-isomorphic.*

Proof. By Corollary 3, $C(I_1)$ and $C(I_2)$ are irreducible. Thus, by Theorem 5 and Schur's Lemma, the following statements are equivalent: (i) I_1 and I_2 are $(KS)_{\alpha^{-1}}$-isomorphic; (ii) $\mathrm{Hom}_{(KS)_{\alpha^{-1}}}(I_1, I_2)$ is K-isomorphic; to K; (iii) $\mathrm{Hom}_{K(G)}(C(I_1), C(I_2))$ is K-isomorphic to K; (iv) $C(I_1)$ and $C(I_2)$ are $K(G)$-isomorphic.

Combining the results of this section, the following theorem is obtained.

THEOREM 6. *Let G be a split extension of H by B. Let T be an irreducible representation of H over an algebraically closed field of characteristic p. Let L be a representation module for T. Let $S = \{b \in B \mid T^{(b)}$ is equiv-*

alent to $T\}$, *and let* $\sigma: b \rightarrow D_b$, *where* D_b *is a linear transformation of* L *such that* $D_b T^{(b)} D_b^{-1} = T$. σ *is a projective representation of* S *with factor set* α. *Let* $(KS)_{\alpha^{-1}}$ *be the* α^{-1}*-twisted group algebra of* S.

Each left ideal I of $(KS)_{\alpha^{-1}}$ determines a left $K(G)$-submodule $C(I)$ of L^G such that

1) $(C(I):K) = (I:K)[B:S](L:K)$.

2) If $S \neq B$, then SH is a proper subgroup of G and there exists a left $K(SH)$-module I' such that $C(I)$ is $K(G)$-isomorphic to $(I')^G$.

3) If $I_1, \cdots, I_t$ are left ideals of $(KS)_{\alpha^{-1}}$ such that $(KS)_{\alpha^{-1}} = I_1 \oplus \cdots \oplus I_t$, then $L^G = C(I_1) \oplus \cdots \oplus C(I_t)$.

4) If I_1 and I_2 are left ideals of $(KS)_{\alpha^{-1}}$, the $\mathrm{Hom}_{(KS)_{\alpha^{-1}}}(I_1, I_2)$ is K-isomorphic to $\mathrm{Hom}_{K(G)}(C(I_1), C(I_2))$.

Let $\mathfrak{T}_I$ denote the representation of G afforded by $C(I)$. Then, if $S \neq B$, $\mathfrak{T}_I$ is an induced representation. If $(KS)_{\alpha^{-1}} = I_1 \oplus \cdots \oplus I_t$, where the I_i are indecomposable left ideals, then $T^G = \mathfrak{T}_{I_1} \oplus \cdots \oplus \mathfrak{T}_{I_t}$ is a decomposition of T^G into indecomposable representations. If $p \nmid [G:1]$, then the $\mathfrak{T}_{I_i}$ are irreducible and $\mathfrak{T}_{I_i}$ and $\mathfrak{T}_{I_j}$ are equivalent if and only if I_i and I_j are $(KS)_{\alpha^{-1}}$-isomorphic.

4. Application. It was shown in § 3 that $C(I)$ is $K(G)$-isomorphic to $(I')^G$ where I' is a left $K(SH)$-module. I' is $K(SH)$-isomorphic to the (inner) tensor product of two modules which afford projective representations of SH. This proves results of Mackey [7] and Clifford [2] when G is a split extension of H by G/H.

I' was defined in § 3 to be the left $K(SH)$-module with basis.

$$\{c(1, s_k, l_\mu) \mid 1 \leqq k \leqq (I:K), 1 \leqq \mu \leqq (L:K)\}$$

where $\{s_k \mid 1 \leqq k \leqq (I:K)\}$ is a basis for I over K and $\{l_\mu \mid 1 \leqq \mu \leqq (L:K)\}$ is a basis for L over K. Define a factor set $(\alpha^{-1})^*$ of SH by

$$(\alpha^{-1})^*(b_1 h_1, b_2 h_2) = \alpha^{-1}(b_1, b_2),$$

for every $b_i \in S$, $h_i \in H$, $i = 1, 2$. Since $S \cap H = \{1\}$, $(\alpha^{-1})^*$ is well-defined. Let $I(SH)$ be the left $(KSH)_{(\alpha^{-1})^*}$-module which is a vector space over K with basis $\{s_k \mid 1 \leqq k \leqq (I:K)\}$ and module operation # such that

$$bh \# s = b \cdot s, \text{ for every } b \in S,\ h \in H,\ s \in I.$$

$I(SH)$ affords the projective representation of SH which is obtained from the projective representation of S afforded by I and the natural homomorphism of SH onto $SH/H = S$.

Define a factor set α^* of SH by

$$\alpha^*(b_1h_1, b_2h_2) = \alpha(b_1, b_2),$$

for every $b_i \in S$, $h_i \in H$, $i = 1, 2$. Let M be the left $(KSH)_{\alpha^*}$-module which is a vector space over K with basis $\{l_\mu \mid 1 \leqq \mu \leqq (L:K)\}$ and such that

$$bhl = D_bT(h)l, \text{ for every } b \in S,\ h \in H,\ l \in L.$$

The tensor product of $I(SH)$ and M has as a basis

$$\{s_k \otimes l_\mu \mid 1 \leqq k \leqq (I:K), 1 \leqq \mu \leqq (L:K)\}.$$

If $b \in S$, $h \in H$, then the action of bh on a basis element of $I(SH) \otimes M$ is given by

$$bh(s_k \otimes l_\mu) = bh \# s_k \otimes bhl_\mu = b \cdot s_k \otimes D_bT(h)l_\mu.$$

$I(SH) \otimes M$ is a left $K(SH)$-module, since $\alpha^*(\alpha^{-1})^* = 1_{SH}$. Let ψ be the K-homomorphism of I' into $I(SH) \otimes M$ such that for every k and μ,

$$\psi\colon c(1, s_k, l_\mu) \to s_k \otimes l_\mu.$$

It is easily checked that ψ is a $K(SH)$-isomorphism of I' onto $I(SH) \otimes M$. Therefore, $C(I)$ is $K(G)$-isomorphic to $(I(SH) \otimes M)^G$. This proves in particular that L^G is $K(G)$-isomorphic to $((KS)_{(\alpha^{-1})^*} \otimes M)^G$.

Suppose $p \nmid [G:1]$, then by the Frobenius Reciprocity Theorem (see [6]), if L' is an irreducible left $K(G)$-module such that L is a component of L' restricted to H, then L' is a component of L^G. Thus, L' is $K(G)$-isomorphic to $(I(SH) \otimes M)^G$, for some I. By Corollary 2, I is irreducible. If I_1 and I_2 are irreducible and $(I_1(SH) \otimes M)^G$ and $(I_2(SH) \otimes M)^G$ are $K(G)$-isomorphic, then, by Corollary 4, I_1 and I_2 are $(KS)_{\alpha^{-1}}$-isomorphic. This proves the following theorem of Mackey in [7]. (Mackey's theorem is more general than the version stated here.)

THEOREM 7. *Let* G *be a finite group with a normal subgroup* H *such that* G *is the split extension of* H *by* G/H. *Let* K *be an algebraically closed field of characteristic* p, $p \nmid [G:1]$. *Let* L *be an irreducible left* $K(H)$*-module and let* H_0 *be the subgroup of* G *consisting of all* $x \in G$ *such that* $x \otimes L$ *is* $K(H)$*-isomorphic to* L. *Then, there exists a factor set* α *for* H_0/H *unique up to equivalence and a factor set* $(\alpha^{-1})^*$ *for* H_0 *and a projective representation* M *of* H_0 *with factor set* $(\alpha^{-1})^*$ *such that* M *restricted to* H *coincides with* L *and such that* $I \to (I(SH) \otimes M)^G$ *sets up a one-to-one correspondence (equiv-*

alent representations identified) between the irreducible projective representations of H_0/H *with factor set* α *and the irreducible representations of* G *such that* L *is a component of their restriction to* H.

Remark. The subgroup H_0 in Theorem 7 corresponds to the subgroup SH.

5. Special cases and some examples. Before Theorem 5 can be applied to a specific group, it is necessary to know the structure of the α^{-1}-twisted group algebra $(KS)_{\alpha^{-1}}$ of S. In general more is known about the structure of the ordinary group algebra $K(S)$, than about $(KS)_{\alpha^{-1}}$. Thus it is of interest to determine some conditions on T, S, and/or H so that $(KS)_{\alpha^{-1}}$ is $K(S)$. In other words, it is desirable to know conditions which imply that $\sigma: b \rightarrow D_b$, where D_b is a linear transformation of L such that $D_b T^{(b)} D_b^{-1} = T$, is equivalent to an ordinary representation of S. Some such conditions are given in the following theorem.

THEOREM 8. *Each of the following conditions implies that* σ *can be taken to be an ordinary representation of* S.

1) $T^{(b)}$ *equivalent to* T *implies* $T^{(b)} = T$;

2) T *is one-dimensional*;

3) *There exists a homomorphism of* S *into* H *such that* $b \rightarrow h_b$ *and* $T(h_b) T^{(b)} T(h_b)^{-1} = T$;

4) *For every* $b \in S$, $\bar{b}: h \rightarrow b^{-1}hb$ *is an inner automorphism* i_{h_b} *of* H, $i_{h_b}: h \rightarrow h_b^{-1} b h_b$, *where the* h_b *are such that* $h_a h_b = h_{ab}$, *for every* $a, b \in S$;

5) H *is a complete group, i.e.,* H *is a group such that all its automorphisms are inner and its center* $= \{1\}$;

6) S *is cyclic.*

Proof of 1). For every $b \in S$, define $D_b = 1_L$, where 1_L denotes the identity transformation on L. Then, $\sigma: b \rightarrow 1_L$ is an ordinary representation of S.

Proof of 2). If T is one-dimensional, then L is one-dimensional. By Schur's Lemma, $\mathrm{Hom}_K(L, L) = K \cdot 1_L$. Thus, D_b, $T^{(b)}(h)$, and $T(h)$ commute for every $h \in H$ and $b \in S$. Therefore, $T^{(b)}$ equivalent to T implies $T^{(b)} = T$ and condition 1) is satisfied.

Proof of 3). For every $b \in S$, define $D_b = T(h_b)$. Then, $D_{ab} = T(h_{ab}) = T(h_a h_b)$, for every $a, b \in S$, since $b \rightarrow h_b$ is a homomorphism. But,

$T(h_a h_b) = T(h_a)T(h_b) = D_a D_b$. Therefore, for every $a, b \in S$, $D_{ab} = D_a D_b$ and $\sigma\colon b \to D_b = T(h_b)$ is an ordinary representation of S.

Proof of 4). The mapping $b \to h_b$ is a homomorphism of S into H. For every $h \in H$,

$$T^{(b)}(h) = T(b^{-1}hb) = T(h_b{}^{-1}hh_b) = T(h_b)^{-1}T(h)T(h_b)$$

so that

$$T(h_b)T^{(b)}(h)T(h_b)^{-1} = T(h).$$

Therefore, $T(h_b)T^{(b)}T(h_b)^{-1} = T$ and condition 3) is satisfied.

Proof of 5). $\bar{b}\colon h \to b^{-1}hb$ is an automorphism of H and since H is complete is therefore an inner automorphism i_{h_b} of H. Then, for every $a, b \in S$ and $h \in H$,

$$h_{ab}{}^{-1}hh_{ab} = b^{-1}a^{-1}hab = b^{-1}h_a{}^{-1}hh_a b = h_b{}^{-1}h_a{}^{-1}hh_a h_b$$

so that $h_a h_b h_{ab}{}^{-1}$ is an element of the center of H. But then, $h_a h_b h_{ab}{}^{-1} = 1$. Therefore, $h_a h_b = h_{ab}$ and condition 4) is satisfied.

Proof of 6). Let $S = [b]$, where b is of order n. It is possible to select D_b such that $(D_b)^n = 1_L$. In order to show this, let D be any non-singular linear transformation of L such that $DT^{(b)}D^{-1} = T$. Then, $D^n T^{(b^n)} D^{-n} = T$. But since $b^n = 1$, $T^{(b^n)} = T$ and, therefore, $D^n T D^{-n} = T$. Thus by Schur's Lemma, $D^n = \zeta \cdot 1_L$, for some $\zeta \in K$. Since K is algebraically closed, there exists a ρ such that $\rho^n = \zeta^{-1}$. Then, $D_b = \rho D$ is such that $D_b T^{(b)} D_b{}^{-1} = T$ and $D_b{}^n = (\rho D)^n = \zeta^{-1}\zeta \cdot 1_L = 1_L$. Define D_{b^i} by $D_{b^i} = (D_b)^i$. Define σ by $\sigma\colon b^i \to D_{b^i} = D_b{}^i$. σ is well-defined since $D_b{}^n = 1_L$, and $b^i = b^j$ implies $i \equiv j \bmod n$. σ is an ordinary representation of S, since

$$D_{b^i}D_{b^j} = D_b{}^i D_b{}^j = D_b{}^{i+j} = D_{b^{i+j}}.$$

This completes the proof of Theorem 8.

Examples of groups H which satisfy one of the conditions of Theorem 8 are: (i) H is abelian and (ii) $H = S_n$, the symmetric group on n symbols, $n \geqq 3$ and $n \neq 6$. For, if H is abelian then all its irreducible representations are one-dimensional so that condition 2) of Theorem 8 is satisfied. If $H = S_n$, $n \geqq 3$ and $n \neq 6$, then H is a complete group (see Kurosh [5, p. 92]), so that condition 5) is satisfied.

Suppose $p \nmid [S\colon 1]$ and $(KS)_{\alpha^{-1}} = K(S)$. Let $\underline{F}^1, \cdots, \underline{F}^t$ be a complete set of inequivalent irreducible matrix representations of S over K and let $\underline{F}^i(b) = (f^i{}_{\mu\nu}(b))$, for every $b \in S$. The functions $\{f^i{}_{\mu\nu} \mid 1 \leqq \mu, \nu \leqq \dim \underline{F}^i\}$ are called the coordinate functions of $\underline{F}^i$. K is assumed to be algebraically closed. Thus, by a theorem of Frobenius and Schur (see [3, § 27]), the set of coordinate functions

$$\{f^{i}{}_{\mu\nu} \mid 1 \leqq \mu, \nu \leqq \dim \underline{F}^{i}, 1 \leqq i \leqq t\}$$

is linearly independent over K.

The elements

$$\{ \sum_{b \in S} f^{i}{}_{\mu\nu}(b^{-1})b \mid 1 \leqq \mu, \nu \leqq \dim \underline{F}^{i}, 1 \leqq i \leqq t\} \tag{5.1}$$

form a basis for $K(S)$. Since for every $b, b' \in S$, $\underline{F}^{i}(bb') = \underline{F}^{i}(b)\underline{F}^{i}(b')$, it can be shown that

$$b'(\sum_{b \in S} f^{i}{}_{\mu\nu}(b^{-1})b) = \sum_{\lambda=1}^{\dim \underline{F}^{i}} f^{i}{}_{\lambda\nu}(b')(\sum_{b \in S} f^{i}{}_{\mu\lambda}(b^{-1})b). \tag{5.2}$$

Thus, for every i and μ, $1 \leqq i \leqq t$ and $1 \leqq \mu \leqq \dim \underline{F}^{i}$, the elements

$$\{ \sum_{b \in S} f^{i}{}_{\mu\lambda}(b^{-1})b \mid 1 \leqq \lambda \leqq \dim \underline{F}^{i}\}$$

form a basis for a left ideal $I_{i,\mu}$ of $(KS)_{\alpha^{-1}}$. Then

$$(KS)_{\alpha^{-1}} = \sum_{i,\mu} \oplus I_{i,\mu} \quad \text{and} \quad L^{G} = \sum_{i,\mu} \oplus C(I_{i,\mu}).$$

$C(I_{i,\mu})$ is of dimension $(I_i : K)[B : S](L : K)$ over K and has as a basis the elements

$$\begin{aligned} \{ \sum_{b \in S} f^{i}{}_{\mu\lambda}(b^{-1})a_{\gamma}b \otimes D_{b}{}^{-1}l_{\eta} \mid 1 \leqq \lambda \leqq \dim \underline{F}^{i}, \\ 1 \leqq \gamma \leqq r, 1 \leqq \eta \leqq (L : K)\}, \end{aligned} \tag{5.3}$$

where $\{l_{\eta} \mid 1 \leqq \eta \leqq (L : K)\}$ is a basis for L over K.

Order the elements (5.3) lexicographically by the triple (γ, λ, η). It will be convenient to consider the basis elements in blocks determined by γ and λ. Thus the $(\gamma - 1)(I_i : K) + \lambda$ block of the basis elements consists of the elements

$$\{ \sum_{b \in S} f^{i}{}_{\mu\lambda}(b^{-1})a_{\gamma}b \otimes D_{b}{}^{-1}l_{\eta} \mid 1 \leqq \eta \leqq (L : K)\}.$$

The matrix representation $\underline{\mathfrak{I}}_{i,\mu}$ of G afforded by $C(I_{i,\mu})$ relative to the basis (5.3) ordered as stated can be determined from (5.2) and the Properties in § 3. $\underline{T}$ is the matrix representation of H afforded by L relative to the basis $\{l_{\eta} \mid 1 \leqq \eta \leqq (L : K)\}$ and $\underline{D}_{b}$ is the matrix of the linear transformation D_{b} of L relative to that basis.

$\underline{\mathfrak{I}}_{i,\mu}(h) =$ (matrix with $\underline{T}^{(a_{\gamma})}(h)$ in the
$((\gamma - 1)(I_i : K) + \lambda, (\gamma - 1)(I_i : K) + \lambda)$ block position;
zeros elsewhere)

$$\begin{aligned}\underline{\mathcal{T}}_{i,\mu}(b) = {} & \text{(matrix with } f^{i}{}_{\lambda\nu}(b')\underline{D}_{b'} \text{ in the} \\ & ((\delta-1)(I_i\colon K)+\lambda,\ (\gamma-1)(I_i\colon K)+\nu) \text{ block position;} \\ & \text{where } ba_\gamma = a_\delta b'\text{; zeros elsewhere)} \\ \underline{\mathcal{T}}_{i,\mu}(a_\beta) = {} & \text{(matrix with } f^{i}{}_{\lambda\nu}(b')\underline{D}_{b'} \text{ in the} \\ & ((\delta-1)(I_i\colon K)+\lambda,\ (\gamma-1)(I_i\colon K)+\nu) \text{ block position;} \\ & \text{where} a_\beta a_\gamma = a_\delta b'\text{; zeros elsewhere).}\end{aligned}$$

From the matrices of $\underline{\mathcal{T}}_{i\mu}$, it follows that $\underline{\mathcal{T}}_{i,\mu} = \underline{\mathcal{T}}_{i,\rho}$, for every $\mu, \rho = 1, \cdots, \dim \underline{F}^i$. $\mathcal{T}_{i,\mu}$ is equivalent to $(\underline{T}_{i,\mu})^G$ where $\underline{\mathcal{T}}_{i,\mu}$ is the matrix representation of SH afforded by $(I_{i,\mu})'$ relative to the basis

$$\{\sum_{b\in S} f^{i}{}_{\mu\lambda}(b^{-1})b \otimes D_b{}^{-1}l_\eta \mid 1 \leqq \nu \leqq \dim \underline{F}^i, 1 \leqq \eta \leqq (L\colon K)\}$$

ordered lexicographically with respect to (ν, η). It will be convenient to consider the basis elements in blocks determined by ν. Then,

$$\begin{aligned}\underline{T}_{i,\mu}(h) = {} & \text{(matrix with } \underline{T}(h) \text{ in the } (\nu,\nu) \text{ block position;} \\ & \text{zeros elsewhere)} \\ \underline{T}_{i,\mu}(\mathrm{b}) = {} & \text{(matrix with } f^{i}{}_{\lambda\nu}(b)\underline{D}_b \text{ in the } (\lambda,\nu) \text{ block position).}\end{aligned}$$

$\underline{T}_{i,\mu}$ is equivalent to the tensor product representation $\underline{I}_{i,\mu}(SH) \otimes \underline{M}$, where $\underline{I}_{i,\mu}(SH)$ and $\underline{M}$ are projective representations of SH such that

$$\begin{aligned}& \underline{I}_{i,\mu}(SH)(h) = 1_{\dim \underline{F}^i} \\ & \underline{I}_{i,\mu}(SH)(b) = \text{(matrix with } f^{i}{}_{\lambda\nu}(b) \text{ in the } (\lambda,\nu) \text{ position)} \\ & \qquad \underline{M}(h) = \underline{T}(h), \text{ and } M(b) = \underline{D}_b.\end{aligned}$$

As an example in which σ cannot be taken to be an ordinary representation, consider the following group G. Let H be the quaternion group of order 8 and let B be the group $[x] \times [y]$, where x and y are of order 2. H is generated by a and b where

$$a^4 = b^4 = 1 \text{ and } a^2 = (ab)^2 = b^2.$$

Define the split extension G of H by B by the automorphisms:

$$\begin{aligned}\bar{x}\colon & a \rightarrow x^{-1}ax = b \\ & b \rightarrow x^{-1}bx = a \\ \bar{y}\colon & a \rightarrow y^{-1}ay = b^3 \\ & b \rightarrow y^{-1}by = a^3.\end{aligned}$$

Let T be the two-dimensional representation of H over the complex field

K. Let L be a representation module of T. Let $\{l_1, l_2\}$ be a basis for L over K such that

$$T(a)l_1 = l_2 \qquad T(a)l_2 = -l_1$$
$$T(b)l_1 = il_1 \qquad T(b)l_2 = -il_2.$$

(See Curtis and Reiner [3, § 47].) Since H has only one two-dimensional representation over K, $T^{(x)}$ and $T^{(y)}$ are equivalent to T. This gives $S = B$.

Define linear transformations D and C on L by

$$Dl_1 = 1/\sqrt{2}(l_1 - il_2) \qquad Dl_2 = 1/\sqrt{2}(il_1 - l_2)$$
$$Cl_1 = 1/\sqrt{2}(l_1 + il_2) \qquad Cl_2 = -1/\sqrt{2}(il_1 + l_2).$$

D and C satisfy the following relationships:

$$DT^{(x)}D^{-1} = T \qquad D^2 = C^2 = 1_L$$
$$CT^{(y)}C^{-1} = T \qquad DC = -CD.$$

Define σ on S by

$$\sigma(x) = D, \sigma(y) = C, \ \sigma(xy) = DC, \text{ and } \sigma(1) = 1_L.$$

Let α denote the factor set of σ. Since $\alpha^{-1} = \alpha$, $(KS)_{\alpha^{-1}} = (KS)_\alpha$.

Let σ' be a projective representation of S with representation space M such that σ' is equivalent to σ. Then, there exists a vector space isomorphism U of M into $\mathrm{Hom}_K(L, L)$ and a function $\rho: S \to K^*$ such that for every $b \in S$,

$$\sigma'(b)U = \rho(b)U\sigma(b).$$

Then, $\sigma'(b) = \rho(b)U\sigma(b)U^{-1}$ and

$$\begin{aligned}\sigma'(x)\sigma'(y) &= (\rho(x)U\sigma(x)U^{-1})(\rho(y)U\sigma(y)U^{-1})\\ &= \rho(x)\rho(y)U\sigma(x)\sigma(y)U^{-1} = -\rho(x)\rho(y)U\sigma(y)\sigma(x)U^{-1}\\ &= -(\rho(y)U\sigma(y)U^{-1})(\rho(x)U\sigma(x)U^{-1}) = -\sigma'(y)\sigma'(x).\end{aligned}$$

Therefore, σ' is not an ordinary representation of S. This proves, in particular, that σ cannot be taken to be an ordinary representation of S, since any other choice for σ is equivalent to σ. Since σ is not equivalent to an ordinary representation, $(KS)_\alpha \neq K(S)$.

The elements $\{1 + y, 1 - y, x + xy, x - xy\}$ form a basis for $(KS)_\alpha$. Let I_1 be the K-subspace of $(KS)_\alpha$ with basis $\{1 + y, x + xy\}$ and let I_2 be the K-subspace with basis $\{1 - y, x - xy\}$. I_1 and I_2 are left ideals of $(KS)_\alpha$.

It is easily checked that the K-homomorphism $\beta: I_1 \to I_2$ defined by

$$\beta(1 + y) = x - xy \text{ and } \beta(x + xy) = 1 - y$$

is a $(KS)_\alpha$-isomorphism of I_1 onto I_2.

I_1 relative to the basis $\{1+y, x+xy\}$ affords a matrix representation $\underline{F}$ given by:

$$\underline{F}(x) = \left\{\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}\right\} \text{ and } \underline{F}(y) = \left\{\begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix}\right\}.$$

If N is a matrix over K such that for every $b \in S$, $N\underline{F}(b) = \underline{F}(b)N$, it can easily be shown that N is a scalar matrix. It then follows from Theorem 2 and the fact that the characteristic of K is zero that I_1 is irreducible. Since I_2 is $(KS)_\alpha$-isomorphic to I_1, I_2 is also irreducible.

By Theorems 4 and 5, $L^G = C(I_1) + C(I_2)$, where $C(I_1)$ is $K(G)$-isomorphic to $C(I_2)$ and $C(I_1)$ and $C(I_2)$ are irreducible. Thus, if T_i denotes the representation of G afforded by $C(I_i)$, $i = 1, 2$, then T_1 and T_2 are equivalent and irreducible. The elements

$$\{(1 \otimes l_1) + (y \otimes Cl_1), (1 \otimes l_2) + (y \otimes Cl_2),$$
$$(x \otimes Dl_1) + (xy \otimes CDl_1), (x \otimes Dl_2) + (xy \otimes CDl_2)\}$$

form a basis for $C(I_1)$. The matrix representation $\underline{T}_1$ of G afforded by $C(I_1)$ relative to this basis is given by:

$$\underline{T}_1(h) = \left\{\begin{matrix} \underline{T}(h) & 0 \\ 0 & \underline{T}(h) \end{matrix}\right\}, \underline{T}_1(x) = \left\{\begin{matrix} 0 & \underline{D} \\ \underline{D} & 0 \end{matrix}\right\}, \underline{T}_1(y) = \left\{\begin{matrix} \underline{C} & 0 \\ 0 & -\underline{C} \end{matrix}\right\}.$$

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

REFERENCES.

[1] E. Artin, C. J. Nesbitt and R. M. Thrall, *Rings With Minimum Condition*, Ann Arbor, 1946.

[2] A. H. Clifford, "Representation induced in an invariant subgroup," *Annals of Mathematics*, vol. 38 (1937), pp. 533-550.

[3] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras*, New York, 1962.

[4] N. Jacobson, *The Theory of Rings*, New York, 1943.

[5] A. G. Kurosh, *The Theory of Groups*, Vol. I, 2nd ed., New York, 1955.

[6] G. W. Mackey, "On induced representations of groups," *American Journal of Mathematics*, vol. 73 (1951), pp. 576-592.

[7] ———, "Unitary representations of group extensions I," *Acta Mathematica*, vol. 99 (1958), pp. 265-311.

ON THE BEHAVIOUR OF THE GEOMETRIC GENUS UNDER GROUND FIELD EXTENSION.*

By PETER L. FALB.[1]

---

1. **The geometric genus.** Let V be an irreducible, normal, r-dimensional variety defined over a ground field k of characteristic $p \neq 0$. We assume that the function field, K, of V over k is separably generated over k. Let $\mathfrak{p}$ be a prime divisor of K/k which is of the first kind on V and let W be the center of $\mathfrak{p}$ on V. Let $u_1,\cdots,u_\rho$ be p-parameters of both W/k and V/k with k_1 as associated field of p-constants ([1]) (i.e., $u_1,\cdots,u_\rho$ are a p-independent set of elements of k such that the ideals of V and W, $\mathfrak{J}(V)$ and $\mathfrak{J}(W)$, have bases consisting of polynomials with coefficients in $k^p(u_1,\cdots,u_\rho)$ and k_1 is a maximal subfield of k, containing k^p, such that $[k_1(u_1,\cdots,u_\rho):k_1]=p^\rho$). For the moment, we shall assume that the p-parameters and field of p-constants have been fixed; and we shall (tacitly) assume that all derivations, which we may consider, are trivial on k_1 (but not necessarily on k). Then, ([2b], p. 33, Theorem 33), we can speak of uniformizing coordinates of $\mathfrak{p}$ relative to the p-parameters $u_1,\cdots,u_\rho$ and the field k_1 of p-constants; i.e., we can consider elements $\xi_1,\cdots,\xi_{\rho+r}$ of the valuation ring $\mathfrak{O}$ of $\mathfrak{p}$ for which there exist derivations D_i of K such that $D_i\xi_j=\delta_{ij}$ $(i,j=1,\cdots,\rho+r)$ and $D_i\mathfrak{O}\subset\mathfrak{O}$ (compare with [1] Definition 1). Uniformizing coordinates $\xi_1,\cdots,\xi_{\rho+r}$ of $\mathfrak{p}$ are characterized by the conditions: (i) $k(\mathfrak{p})$ is separable algebraic over $k_1(\bar{\xi}_1,\cdots,\bar{\xi}_{\rho+r})$ where bar denotes $\mathfrak{M}$-residue, $\mathfrak{M}$ being the maximal ideal of $\mathfrak{O}$ and (ii) $k_1[\xi_1,\cdots,\xi_{\rho+r}]$ contains a set of uniformizing parameters of $\mathfrak{p}$ ([1] Theorem 1).

Since the space $\mathfrak{D}(K/k_1)$ of derivations of K over k_1 has dimension $\rho+r$, we may speak of differentials of K, of degree $\rho+r$, and for any $\rho+r+1$ elements A, $z_1,\cdots,z_{\rho+r}$ of K, we have a differential $A\,dz_1\cdots dz_{\rho+r}$ of K. Furthermore, if $\xi_1,\cdots,\xi_{\rho+r}$ are uniformizing coordinates of $\mathfrak{p}$, then every $\rho+r$-fold differential of K can be written in the form $A\,d\xi_1\cdots d\xi_{\rho+r}$, $A\in K$.

Definition 1. *If $x_1,\cdots,x_r$ is a separating transcendence basis of K/k, if $\xi_1,\cdots,\xi_{\rho+r}$ are uniformizing coordinates of $\mathfrak{p}$, and if $dx_1\cdots dx_r\,du_1\cdots du_\rho$ $=A\,d\xi_1\cdots d\xi_{\rho+r}$ (where the u_i are the given p-parameters), then the order*

---

* Received December 13, 1961.

[1] Lincoln Laboratory, Massachusetts Institute of Technology. Operated with the joint support of the U. S. Army, Navy and Air Force.

of $dx_1 \cdots dx_r$ *at* $\mathfrak{p}$ *(in symbols:* $v_{\mathfrak{p}}(dx_1 \cdots dx_r)$*) is the integer* $v_{\mathfrak{p}}(A)$ *where* $v_{\mathfrak{p}}$ *is the valuation defined by* $\mathfrak{p}$.

PROPOSITION 1. *Let* $x_1, \cdots, x_r$ *be a separating transcendence basis of* K/k *such that* $\mathfrak{p}$ *is of the first kind with respect to* $R = k[x_1, \cdots, x_r]$ *(such a transcendence basis exists since* $\mathfrak{p}$ *is of the first kind on* V*). Let* R_x *be the integral closure of* R *in* K *and let* $\mathfrak{D}_x = \mathfrak{D}_{R_x/k[x]}$ *be the different of* K *over* $k(x_1, \cdots, x_r)$*. Then* $v_\rho(dx_1 \cdots dx_r) = v_{\mathfrak{p}}(\mathfrak{D}_x)$.

Proof. For $r = 1$, this is the main result of [1]. Let $\tilde{k} = k(x_1, \cdots, x_{r-1})$ where we assume (since $\mathfrak{p}$ is of the first kind on R) that $\mathfrak{p}$ is trivial on $\tilde{k}$. If $\tilde{k}_1 = k_1(x_1{}^p, \cdots, x_{r-1}{}^p)$, then $[\tilde{k} : \tilde{k}_1] = p^{\rho+r-1}$. Let $M = k[x_1, \cdots, x_{r-1}] - (0)$. Then we have ([2a] p. 299, Theorem 27) $v_{\mathfrak{p}}(\mathfrak{D}_{R_{xM}/k[x_r]}) = v_{\mathfrak{p}}(\mathfrak{D}_x)$; and since $\mathfrak{p}$ is a prime divisor of $K/\tilde{k}$, $v_{\mathfrak{p}}(\mathfrak{D}_x) = v_{\mathfrak{p}}(\bar{d}x_r)$ where $\bar{d}x_r$ is the differential of x_r over $\tilde{k}$. But $v_{\mathfrak{p}}(\bar{d}x_r) = v_{\mathfrak{p}}(dx_r dx_1 \cdots dx_{r-1} du_1 \cdots du_\rho)$ since $u_1, \cdots, u_\rho, x_1, \cdots, x_{r-1}$ are p-parameters for $\mathfrak{p}$ over $\tilde{k}$. Thus we see that $v_{\mathfrak{p}}(\mathfrak{D}_x) = v_{\mathfrak{p}}(\bar{d}x_r) = v_{\mathfrak{p}}(dx_1 \cdots dx_r)$.

COROLLARY. $v_{\mathfrak{p}}(dx_1 \cdots dx_r)$ *is well defined.*

Definition 2. *If* ω *is an* r*-fold differential of* K *over* k *and if* $\omega = A\,dx_1 \cdots dx_r$ *where* $\{x_1, \cdots, x_r\}$ *is a separating transcendence basis of* K/k*, then the order of* ω *at* $\mathfrak{p}$*,* $v_{\mathfrak{p}}(\omega)$*, is the integer* $v_{\mathfrak{p}}(A) + v_{\mathfrak{p}}(dx_1 \cdots dx_r)$ *and* ω *is regular at* $\mathfrak{p}$ *if* $v_{\mathfrak{p}}(\omega) \geqq 0$.

Let $\Omega(V/k) = \{\omega : \omega$ is an r-fold differential of K/k and ω is regular at all prime divisors $\mathfrak{p}$ of K/k which are of the first kind on $V\}$. It is clear that $\Omega(V/k)$ is a vector space over k. Let $\tilde{\mathfrak{p}}$ be any prime divisor of K/k and let ω be a differential of K/k. Then, ([2b], p. 89, Theorem 31 and p. 97), there is a normal model $\tilde{V}$ of K/k on which $\tilde{\mathfrak{p}}$ is of the first kind and consequently, $v_{\tilde{\mathfrak{p}}}(\omega)$, is defined (relative to the model $\tilde{V}$). We assert that $v_{\tilde{\mathfrak{p}}}(\omega)$ is independent of the choice of model $\tilde{V}$. To see this, we observe that if $\tilde{V}$, $\tilde{V}_1$ are two normal models, with respect to which $\mathfrak{p}$ is of the first kind, then there is a set $z_1, \cdots, z_\rho$ of p-parameters and associated field $\mathfrak{k}_1$ of p-constants for $\tilde{V}$, $\tilde{V}_1$, $\bar{W}$, $\bar{W}_1$ where $\bar{W}$, $\tilde{W}_1$ are the centers of $\tilde{\mathfrak{p}}$ on V, V_1 respectively,[2] and that there exists a set of uniformizing coordinates $\xi_1, \cdots, \xi_{\rho+r}$ of both the center $\tilde{W}$ of $\tilde{\mathfrak{p}}$ on $\tilde{V}$ and the center $\tilde{W}_1$ of $\tilde{\mathfrak{p}}_1$ and $\tilde{V}_1$. Our assertion follows immediately. It makes sense therefore to speak of the regularity of a differential of K/k at *any* prime divisor of K/k. Let $\Omega(K/k) = \{\omega : \omega$ is an r-fold differential of K/k and ω is regular at all prime divisors of $K/k\}$. It is clear that $\Omega(K/k)$ is a vector space over k and that $\Omega(K/k) \subset \Omega(V/k)$. We now assume that K/k is a regular extension.

---

[2] Compare with [1] Section 1.

PROPOSITION 2. *Let $p_g(V/k) = \dim_k \Omega(V/k)$. Then $p_g(V/k)$ is finite.*

Proof. Let $x_1, \cdots, x_r$ be a separating transcendence basis of K/k and let $\eta_1, \cdots, \eta_{\rho+r}$ be $x_1, \cdots, x_r, u_1, \cdots, u_\rho$ (where the u_i are our p-parameters). Suppose that $k[V] = k[x_1, \cdots, x_N]$ is the coordinate ring of V. We consider the set $\{D_i x_j\}$ $(i = 1, \cdots, \rho + r,\ j = r + 1, \cdots, N)$ where the D_i are the derivations of K (over the field k_1 of p-constants) such that $D_i \eta_s = \delta_{is}$. Then there are only a finite number of prime divisors of the first kind in V which are polar divisors of some $D_i x_j$. It follows that if $\mathfrak{p}$ is any other prime divisor of K/k (of the first kind on V), then $\eta_1, \cdots, \eta_{\rho+r}$ are uniformizing coordinates of $\mathfrak{p}$ (i.e., satisfy the conditions of [1] Theorem 1) since $\mathfrak{O}_\mathfrak{p}$ is a quotient ring of $k[x_1, \cdots, x_N]$. Let $\mathfrak{P} = \{\mathfrak{p}_1, \cdots, \mathfrak{p}_n\}$ be the set of prime divisors of the first kind on V at which the η_i are *not* uniformizing coordinates. Let $\xi_1, \cdots, \xi_{\sigma+r}$ be uniformizing coordinates of $\mathfrak{p}_1$ (relative to p-parameters $u_1, \cdots, u_\rho,\ u_{\rho+1}, \cdots, u_\sigma$ i.e., $\{u_1, \cdots, u_\rho, u_{\rho+1}, \cdots, u_\sigma\}$ is a set of p-parameters for both $\mathfrak{p}_1$ and V. Then $d\eta_1 \cdots d\eta_{\rho+r} du_{\rho+1} \cdots du_\sigma = A\, d\xi_1 \cdots d\xi_{\sigma+r}$ and it follows that either $v_{\mathfrak{p}_1}(\eta_i) < 0$ for some i or $\mathfrak{p}_1$ is a component of the divisor of A. Let $\omega \in \Omega(V/k)$ and suppose that

$$\omega du_1 \cdots du_\rho du_{\rho+1} \cdots du_\sigma = B\, d\eta_1 \cdots d\eta_{\rho+r} du_{\rho+1} \cdots du_\sigma = BA\, d\xi_1 \cdots d\xi_{\sigma+r}.$$

Then, $v_{\mathfrak{p}_1}(\omega) = v_{\mathfrak{p}_1}(B) + v_{\mathfrak{p}_1}(A) \geqq 0$. Let $s(\mathfrak{p}_1) = \min\{-v_{\mathfrak{p}_1}(A), 0\}$. We define $s(\mathfrak{p}_2), \cdots, s(\mathfrak{p}_n)$ in a similar way and we set $Z = -\sum s(\mathfrak{p}_i)\mathfrak{p}_i$. Then, if $\omega du_1 \cdots du_\rho = a\, d\eta_1 \cdots d\eta_{\rho+r}$, we have, (a), (the divisor of a), is such that $(a) + Z \geqq 0$. It follows that $\Omega(V/k) \subset \{f : (f) + Z \geqq 0\}$ and therefore, ([2b] p. 357), since K/k is a regular extension, we have $p_g(V/k) < \infty$.

COROLLARY. *If $p_g(K/k) = \dim_k \Omega(K/k)$, then $p_g(K/k)$ is finite.*

Definition 3. The geometric genus of V/k is the integer $p_g(V/k)$ and the geometric genus of K/k is the integer $p_g(K/k)$.

Let k' be an extension field of k, which is linearly disjoint from K over k and let K' be obtained from K by extending the constant field to k'. We observe that K'/k' is again a regular extension so that we may speak of the geometric genus of K'/k'. We shall investigate the relationship between $p_g(K/k)$ and $p_g(K'/k')$ in the sequel.

2. The main result for finitely generated extensions.

We now assume that k' is a finitely generated extension extension of k. The main object of this section is the proof of the following theorem which is well-known in the case of curves:

THEOREM 1. *If K/k is a regular extension, if k' is a finitely generated extension of k which is linearly disjoint from K over k, and if K'/k' is obtained from K/k by extending the ground field from k to k', then $p_g(K'/k') \leqq p_g(K/k)$ with equality if k' is a separably generated extension of k (equivalently, if k' is a separable extension of k).*

Before proceeding to prove this theorem, we shall make several remarks which will facilitate our exposition of the proof.

Remark 1. It will suffice to prove the theorem in three cases, namely: (i) k' is a purely transcendental extension of k, (ii) k' is a separable algebraic extension of k, (iii) k' is a purely inseparable extension of k. Furthermore, we need only demonstrate the theorem for simple extensions in each of these cases as once we have done this an easy induction on the number of generators of k' over k will give us the main result.

Remark 2. If ω is a differential of K/k and $\omega = A dx_1 \cdots dx_r$ where $\{x_1, \cdots, x_r\}$ is a separating transcendence basis of K/k (a fortiori of K'/k'), then $\omega' = A d'x_1 \cdots d'x_r$ is a well-defined differential of K'/k' which we call the *cotrace* of ω and which we denote by $\text{cotr}\,\omega$.

Remark 3. Let $\mathfrak{p}$ be a prime divisor of K/k and let $\mathfrak{p}'$ be a prime divisor of K'/k' which lies over $\mathfrak{p}$. Let $u_1, \cdots, u_\rho$ be p-parameters of $\mathfrak{p}$ with k_1 as a field of associated p-constants. Then, if $k' = k(\alpha)$ where α is transcendental over k, $u_1, \cdots, u_\rho, \alpha$ are p-parameters of $\mathfrak{p}'$ (which is unique) with $k_1(\alpha^p)$ as an associated field of p-constants. This is an immediate consequence of the well-known facts that if R is the coordinate ring of a (normal) model of K/k, then $k'[R]$ is the coordinate ring of a (normal) model of K'/k'; that if $\mathfrak{A}$ is a prime ideal in R, then $k'[R] \cdot \mathfrak{A}$ is a prime ideal in $k'[R]$; and that prime divisors of a function field correspond to minimal prime ideals in the coordinate ring of some normal model of the function field ([2b], pp. 88-99).

Remark 4. Let $\mathfrak{p}$, $\mathfrak{p}'$, $\{u_1, \cdots, u_\rho\}$ and k_1 be as in the previous remark and let $k' = k(\alpha)$ where α is separable algebraic over k. We may choose a set $\{u_1, \cdots, u_\rho, u_{\rho+1}, \cdots, u_\sigma\}$ of p-parameters for $\mathfrak{p}'$ where the u_i are all in k. To see this, we note that if $\mathfrak{A}$ is an ideal in $k[\alpha][X_1, \cdots, X_n]$ (observe that $k' = k(\alpha) = k[\alpha]$ as α is algebraic over k), then if $\{f_\nu\}$ is a basis of $\mathfrak{A}$ where we view the f_ν as elements of $k[\alpha, X_1, \cdots, X_n]$, we can find p-independent elements $u_1, \cdots, u_\rho, u_{\rho+1}, \cdots, u_\sigma$ in k such that the f_ν have coefficients in $k^p(u_1, \cdots, u_\sigma)(\alpha)$ and since α is separable algebraic over k, we may suppose α is separable algebraic over $k^p(u_1, \cdots, u_\sigma)$. It follows that $\alpha \in k^p(\alpha^p)(u_1, \cdots, u_\sigma)$ and that $u_1, \cdots, u_\sigma$ are p-independent over $k^p(\alpha^p)$ (since the derivations D_i^0 of $k^p(u_1, \cdots, u_\sigma)$ such that $D_i^0 u_j = \delta_{ij}$ can be extended to be derivations over k'^p.) Our assertion follows (where $\mathfrak{A}$ is the ideal of the center of $\mathfrak{p}'$ on a suitable model). If k_2 is a field of p-constants associated with $u_1, \cdots, u_\sigma$ as p-parameters of $\mathfrak{p}$, then $k_2(\alpha^p)$ is a field of p-constants associated with $u_1, \cdots, u_\sigma$ relative to $\mathfrak{p}'$.

Remark 5. Let $\mathfrak{p}$, $\mathfrak{p}'$ be as in the two previous remarks and let $k' = k(\alpha)$ be a purely inseparable extension of degree p. Then we may choose a set $\{u_1, \cdots, u_\sigma, \alpha^p\}$ of p-parameters of $\mathfrak{p}$ such that $\{u_1, \cdots, u_\sigma, \alpha\}$ are p-parameters of $\mathfrak{p}'$. To see this, we observe that by the argument of the previous remark, we can find p-independent elements $u_1, \cdots, u_\sigma, \alpha^p$ which are p-parameters of $\mathfrak{p}$ such that the coefficients of the polynomials which form a basis of the ideal of the center of $\mathfrak{p}'$ on a suitable model are in $k^p(u_1, \cdots, u_\sigma, \alpha^p)(\alpha) = k^p(u_1, \cdots, u_\sigma, \alpha)$. It is clear that $u_1, \cdots, u_\sigma, \alpha$ are p-independent. If k_1 is a field of p-constants associated with $\{u_1, \cdots, u_\sigma, \alpha^p\}$, then $k_2 = k_1(\alpha^p)$ is a field of p-constants associated with $\{u_1, \cdots, u_\sigma, \alpha\}$.

Remark 6. If $\mathfrak{p}$ is a prime divisor of K/k and if $A, x_1, \cdots, x_r$ are in $\mathfrak{O}_\mathfrak{p}$, then $A dx_1 \cdots dx_r$ is regular at $\mathfrak{p}$. In fact, if $t_1, \cdots, t_{\rho+r}, B$ are elements of $\mathfrak{O}_\mathfrak{p}$ (where we have a set $\{u_1, \cdots, u_\rho\}$ of p-parameters of $\mathfrak{p}$), then $B dt_1 \cdots dt_{\rho+r}$ is also regular at $\mathfrak{p}$ in the sense that if $B dt_1 \cdots dt_{\rho+r} = C d\xi_1 \cdots d\xi_{\rho+r}$ where $\xi_1, \cdots, \xi_{\rho+r}$ are uniformizing coordinates of $\mathfrak{p}$, then $C \in \mathfrak{O}_\mathfrak{p}$. This follows immediately from the fact that the derivations D_i such that $D_i \xi_j = \delta_{ij}$ map $\mathfrak{O}_\mathfrak{p}$ into $\mathfrak{O}_\mathfrak{p}$.

We now proceed to prove the theorem in each of the three cases where we assume that $k' = k(\alpha)$ is a simple extension of k. Each case will be treated in a separate proposition.

PROPOSITION 3. *If α is transcendental over k, then $p_g(K/k) = p_g(K'/k')$.*

Proof. We first show that $p_g(K/k) \leqq p_g(K'/k')$. It will suffice to prove that if $\omega \in \Omega(K/k)$, then $\omega' = \operatorname{cotr}\omega \in \Omega(K'/k')$ as it is clear that for any differentials $\omega_1, \cdots, \omega_t$ in $\Omega = \Omega(K/k)$ which are k-independent and for which $\omega'_i = \operatorname{cotr}\omega_i$ are in $\Omega' = \Omega(K'/k')$, we have the k'-independence of the ω'_i. Let $\mathfrak{p}'$ be a prime divisor of K'/k'. If $\mathfrak{p}'$ is trivial on K, we conclude from Remark 6 that ω' is regular at $\mathfrak{p}'$. On the other hand, if $\mathfrak{p}'$ is not trivial on K, then $\mathfrak{p}'$ lies over a prime divisor $\mathfrak{p}$ of K/k; and if we choose

p-parameters $u_1, \cdots, u_\rho$ as in Remark 3 and uniformizing coordinates $\xi_1, \cdots, \xi_{\rho+r}$ of $\mathfrak{p}$ relative to the u_i, then $\omega du_1 \cdots du_\rho = A d\xi_1 \cdots d\xi_{\rho+r}$ with $A \in \mathfrak{O}_\mathfrak{p}$ will imply that $\omega' du_1 \cdots du\, d\alpha = A d\xi_1 \cdots d\xi_{\rho+r} d\alpha$ and hence, we have, by Remark 6, the regularity of ω' at $\mathfrak{p}'$. In any case, ω' is regular at $\mathfrak{p}'$ for any prime divisor $\mathfrak{p}'$ of K'/k' and our assertion follows.

Now let $\tilde{\omega} \in \Omega'$ and observe that there exists an ω, which is a differential of K/k, such that $\tilde{\omega} = A' \operatorname{cotr} \omega$, $A' \in K'$. Suppose that $A' = f(\alpha)/g(\alpha)$, $f, g \in K[\alpha]$ which are relatively prime and let $\mathfrak{p}'$ be a prime divisor of K'/k' which is trivial on K. We observe that $v_{\mathfrak{p}'}(\operatorname{cotr} \omega) = 0$ since we can choose p-parameters $u_1, \cdots, u_\sigma$, α of $\mathfrak{p}'$ with the u_i in k[3] and we can extend the derivations $D_{x_1}, \cdots, D_{x_r}$ of K to K', where $x_1, \cdots, x_r$ is a separating transcendence basis of K/k, by setting $D_{x_i}\alpha = 0$; i.e., we have $x_1, \cdots, x_r, u_1, \cdots, u_\sigma, \alpha$ as uniformizing coordinates at $\mathfrak{p}'$ from which our assertion follows immediately. Now $g(\alpha) = c \prod g_i(\alpha)$ where $c \in K$ and the g_i are irreducible elements of $K[\alpha]$ which either generate prime divisors $\mathfrak{p}'_i$ of K'/k' trivial on K or are in k$'$. Suppose, for example, that g_1 generates $\mathfrak{p}'_1$; then $v_{\mathfrak{p}'_1}(g_1) \geqq 1$ implies that $v_{\mathfrak{p}'_1}(\tilde{\omega}) \leqq -1$ which contradicts the regularity of $\tilde{\omega}$. It follows that $A' = \sum_{j \geqq 0} (a_j/b)\alpha^j$ where $a_j \in K$ and $b \in k'$. Let $\omega_j = a_j\omega$. We claim $\omega_j \in \Omega$ for all j. Suppose $\mathfrak{p}$ is a prime divisor of K/k and let $\mathfrak{p}'$ be the prime divisor of K'/k' lying over $\mathfrak{p}$. We choose p-parameters $u_1, \cdots, u_\rho$ as in Remark 3 and we let $\xi_1, \cdots, \xi_{\rho+r}$ be uniformizing coordinates of $\mathfrak{p}$ relative to the u_i. Then $\omega du_1 \cdots du_\rho = A d\xi_1 \cdots d\xi_{\rho+r}$ and

$$\tilde{\omega} du_1 \cdots du_\rho d\alpha = \sum \{ (\frac{a_j}{b} \operatorname{cotr} \omega \, du_1 \cdots du_\rho d\alpha) \cdot \alpha^j \}$$
$$= \sum (\frac{a_j A}{b} d\xi_1 \cdots d\xi_{\rho+r} d\alpha) \alpha^j.$$

It follows that $\frac{1}{b} \sum a_j A \alpha^j$ is in $\mathfrak{O}_{\mathfrak{p}'}$. But $b \in k'$ is a unit in $\mathfrak{O}_{\mathfrak{p}'}$ and thus we have $\sum (a_j A)\alpha^j \in \mathfrak{O}_{\mathfrak{p}'} \cap K[\alpha] = \mathfrak{O}_\mathfrak{p}[\alpha]$.[4] Consequently, $(a_j A)$ is in $\mathfrak{O}_\mathfrak{p}$ for all j and ω_j is therefore regular at $\mathfrak{p}$. This completes the proof of our proposition.

---

[3] To see this, we let $\mathfrak{A}$ be an ideal in $k'[X_1, \cdots, X_n]$ and $\{f_\nu\}$ be a basis of $\mathfrak{A}$. Then $f_\nu = \sum a_{\nu i}(\alpha, X_1, \cdots, X_n)/b(\alpha)$ where $a_{\nu i}(\alpha, X_1, \cdots, X_n) \in k[\alpha, X_1, \cdots, X_n]$ $b(\alpha) \in k[\alpha]$ and we consider the field obtained by adjoining coefficients of $a_{\nu i}$ and b and also the element α to k^p.

[4] Let $h(\alpha) \in \mathfrak{O}_{\mathfrak{p}'} \cap K[\alpha]$. Then $h(\alpha) = \sum (c_i/c)\alpha^i$ with c_i, c in $\mathfrak{O}_\mathfrak{p}$. Let c_{i_0} be such that $v_\mathfrak{p}(c_{i_0}) = \min\{v_\mathfrak{p}(c_i)\}$. Then $h(\alpha) = (c_{i_0}/c) h_1(\alpha)$ where $h_1(\alpha) \in \mathfrak{O}_\mathfrak{p}[\alpha]$ which has one coefficient equal to 1. If $h(\alpha)$ were not in $\mathfrak{O}_\mathfrak{p}[\alpha]$, then, applying the place determined by $\mathfrak{p}'$ to $h(\alpha) \cdot c/c_{i_0} = h_1(\alpha)$, we would obtain the contradiction $0 = 1$. Hence $h(\alpha) \in \mathfrak{O}_\mathfrak{p}[\alpha]$.

PROPOSITION 4. *If α is separable algebraic over k, then $p_g(K/k) = p_g(K'/k')$.*

Proof. Let $\omega \in \Omega(K/k)$ and let $\omega' = \mathrm{cotr}\omega$. Let $\mathfrak{p}'$ be a prime divisor of K'/k' and suppose that $\mathfrak{p}'$ lies over $\mathfrak{p}$. Then we choose p-parameters $u_1, \cdots, u_\sigma$ of $\mathfrak{p}'$ as in Remark 4 and it will follow that if $\xi_1, \cdots, \xi_{\sigma+r}$ are uniformizing coordinates of $\mathfrak{p}$ (relative to the u_i), then $\xi_1, \cdots, \xi_{\sigma+r}$ are uniformizing coordinates of $\mathfrak{p}'$. It follows immediately that the regularity of ω at $\mathfrak{p}$ implies the regularity of ω' at $\mathfrak{p}'$. Then arguing as in the beginning of the proof of Proposition 3, we see that $p_g(K/k) \leqq p_g(K'/k')$.

On the other hand, suppose that $\bar{\omega}$ is in $\Omega(K'/k')$ and observe that there exists a differential ω of K/k such that $\bar{\omega} = A' \,cotr\omega$, $A' \in K'$. Then $\bar{\omega} = \sum_{j=0}^{[k':k]-1} (a_j \mathrm{cotr}\omega)\alpha^j$ where $a_j \in K$ and therefore, it will suffice to show that $\omega_j = a_j\omega$ are in $\Omega(K/k)$ (compare with the previous proof). Let $\mathfrak{p}$ be a prime divisor of K/k and let $\mathfrak{p}'$ be a prime divisor of K'/k' lying over $\mathfrak{p}$. Let $\xi_1, \cdots, \xi_{\sigma+r}$ be uniformizing coordinates at $\mathfrak{p}$ relative to p-parameters $u_1, \cdots, u_\sigma$ chosen as in the first half of the proof. Suppose $a_j\omega du_1 \cdots du_\sigma = a_j A d\xi_1 \cdots d\xi_{\sigma+r}$. Then we have $\sum(a_j A)\alpha^j \in \mathfrak{O}_{\mathfrak{p}'}$ as $\bar{\omega}$ is regular at $\mathfrak{p}'$ and the ξ_i are also uniformizing coordinates at $\mathfrak{p}'$. It follows that $\sum(a_j A)\alpha^j \in \cap \mathfrak{O}_{\mathfrak{p}'}$ where $\mathfrak{p}'$ runs through the set of prime divisors of K'/k' lying over $\mathfrak{p}$. It is well-known that $\cap \mathfrak{O}_{\mathfrak{p}'} = \mathfrak{O}_{\mathfrak{p}}[\alpha]$ and therefore, we have $a_j A \in \mathfrak{O}_{\mathfrak{p}}$ and so ω_j is regular at $\mathfrak{p}$. This completes the proof of our proposition.

PROPOSITION 5. *If α is purely inseparable over k, then $p_g(K'/k') \leqq p_g(K/k)$.*

Proof. We observe first of all that it will be sufficient to prove the proposition under the assumption that $[k':k] = p$ as once the result is established in this case, it will follow in general by a simple induction. We let $\bar{\omega}$ be in $\Omega(K/k')$. Then there exist differentials ω_j of K over k such that $\bar{\omega} = \sum_{j=0}^{p-1}(\mathrm{cotr}\omega_j)\alpha^j$ and it will suffice to show that the ω_j are regular differentials of K/k (compare with the previous propositions). Let $\mathfrak{p}$ be a prime divisor of K/k and let $\mathfrak{p}'$ be its unique extension to a prime divisor of K'/k'. We have two cases: (a) $k'(\mathfrak{p}')$ is a purely inseparable extension of $k(\mathfrak{p})$ of degree p; (b) $e_{\mathfrak{p}'/\mathfrak{p}} = p$ and $k(\mathfrak{p}') = k(\mathfrak{p})$. We let $u_1, \cdots, u_\sigma, \alpha^p$ be p-parameters of $\mathfrak{p}$ chosen as in Remark 5.

Case (a). Let $\eta_1, \cdots, \eta_{\sigma+r+1}$ be uniformizing coordinates of $\mathfrak{p}$ relative to

$\{u_i, \alpha^p\}$ with (say) η_1 a uniformizing parameter of $\mathfrak{p}$ (a fortiori of $\mathfrak{p}'$).[5] Then we assert that (say) $\eta_1, \cdots, \eta_{\sigma+r}, \alpha$ are uniformizing coordinates at $\mathfrak{p}'$ relative to $\{u_i, \alpha\}$. Suppose, for the moment that this has been established and let $\mathfrak{O} = \mathfrak{O}_{\mathfrak{p}}$, $\mathfrak{O}' = \mathfrak{O}_{\mathfrak{p}'}$. Then $\mathfrak{O}' = \mathfrak{O} + \mathfrak{O}\alpha + \cdots + \mathfrak{O}\alpha^{p-1}$ and

$$\tilde{\omega} du_1 \cdots du_\sigma d\alpha = \sum_{j=0}^{p-1} (a_j d'x_1 \cdots d'x_r du_1 \cdots du_\sigma d\alpha) \alpha^j$$

where $x_1, \cdots, x_r$ is a separating transcendence basis of K/k (a fortiori of K'/k'). Let $D_1, \cdots, D_{\sigma+r+1}$ be derivations of K such that $D_i \eta_j = \delta_{ij}$ and denote by $D'_1, \cdots, D'_{\sigma+r}$ the extensions of $D_1, \cdots, D_{\sigma+r}$ to K' and let D'_α be the derivation of K'/k' with respect to α. It follows that, if

$$\delta = (d'x_1 \cdots d'x_r du_1 \cdots du_\sigma)(D'_1, \cdots, D'_{\sigma+r}),$$

then:

$$(\tilde{\omega} du_1 \cdots du_\sigma d\alpha)(D'_1, \cdots, D'_{\sigma+r}, D'_\alpha) = \sum_{j=0}^{p-1} (a_j \delta) \alpha^j.$$

However,

$$(\omega_j du_1 \cdots du_\sigma d\alpha^p)(D_1, \cdots, D_{\sigma+r}, D_{\sigma+r+1}) = a_j (dx_1 \cdots dx_r du_1 \cdots du_\sigma d\alpha^p)$$

$(D_1, \cdots, D_{\sigma+r}, D_{\sigma+r+1}) = a_j \delta D_{\sigma+r+1} \alpha^p$ since $D_i \alpha^p = 0$, $i = 1, \cdots, \sigma + r$. Since we have $\sum_{j=0}^{p-1} (a_j \delta) \alpha^j \in \mathfrak{O}'$ and $D_{\sigma+r+1} \alpha^p \in \mathfrak{O}$, we know that $a_j \delta D_{\sigma+r+1} \alpha^p$ is in $\mathfrak{O}$ and consequently, that ω_j is regular at $\mathfrak{p}$. It remains to show that our assertion about the uniformizing coordinates is true. We let k_1, k_2 be fields of p-constants as in Remark 5. Then, ([1] Theorem 1), it will be enough to show that $k'(\mathfrak{p}')$ is separable algebraic over $k_2(\bar{\eta}_2, \cdots, \bar{\eta}_{\sigma+r}, \alpha)$ (say) since η_1 is a uniformizing parameter of $\mathfrak{p}$ and $\mathfrak{p}'$. However, if $\mathcal{D}(\mathfrak{p}'/k_2)$ is the space of derivations of $k'(\mathfrak{p}')$ over k_2, then, ([1] Lemma 1), $\dim \mathcal{D}(\mathfrak{p}'/k_2) \leqq \sigma + r$. On the other hand, $k'(\mathfrak{p}')$ is separable algebraic over $k_2(\bar{\eta}_2, \cdots, \bar{\eta}_{\sigma+r+1}, \alpha) = L$ from which it follows that $\dim \mathcal{D}(L/k_2) \leqq \sigma + r$ and therefore, $[L : k_2(L^p)] \leqq p^{\sigma+r}$. As $\alpha \notin k_2(L^p)$, some $\bar{\eta}_i$, say, $\bar{\eta}_{\sigma+r+1}$ is p-dependent on $\bar{\eta}_2, \cdots, \bar{\eta}_{\sigma+r}, \alpha$ and our assertion follows.

Case (b). We let $\eta_0, \cdots, \eta_{\sigma+r}$ be uniformizing coordinates at $\mathfrak{p}$ (relative to $\{u_i, \alpha^p\}$ with η_0 a uniformizing parameter at $\mathfrak{p}$, $\eta_0 = t^p$ where t is a uniformizing parameter of $\mathfrak{p}'$. Then, if $\mathfrak{O}' = \mathfrak{O}_{\mathfrak{p}'}$ and $\mathfrak{O} = \mathfrak{O}_{\mathfrak{p}}$, we have

$$\mathfrak{O}' = \mathfrak{O} + \mathfrak{O}t + \cdots + \mathfrak{O}t^{p-1}.$$

Let $D_0, \cdots, D_{\sigma+r}$ be derivations of K such that $D_i \eta_j = \delta_{ij}$ and let $D'_1, \cdots, D'_{\sigma+r}$ be their extensions to K'. We let D'_t be the derivation of K'/k' with respect to t and we observe that $t, \eta_1, \cdots, \eta_{\sigma+r}$ are uniformizing coordinates of $\mathfrak{p}'$. ([1] Theorem 1).

---

[5] ([1] Corollary of Theorem 1).

Now $\tilde{\omega} du_1 \cdots du_\sigma d\alpha = \sum_{j=0}^{p-1} (a_j d'x_1 \cdots d'x_r du_1 \cdots du_\sigma d\alpha)\alpha^j$ where $x_1, \cdots, x_r$ is a separating transcendence basis of K/k such that $x_i \in \mathfrak{O}$, $i = 1, \cdots, r$. Since $D'_t x_1 = \cdots = D'_t x_r = D'_t u_1 = \cdots = D'_t u_\sigma = 0$ we have, if

$$\delta = (d'x_1 \cdots d'x_r\, du_1 \cdots du_\sigma)(D'_1, \cdots, D'_{\sigma+r}),$$

$$(\tilde{\omega} du_1 \cdots du_\sigma d\alpha)(D'_1, \cdots, D_{\sigma+r}, D'_t) = \sum_{j=0}^{p-1} (a_j\delta)(D'_t\alpha)\alpha^j.$$

Furthermore, since $x_1, \cdots, x_r, u_1, \cdots, u_\sigma$ are in K, we have

$$(a_j dx_1 \cdots dx_r \cdots du_\sigma d\alpha^p)(D_1, \cdots, D_{\sigma+r}, D_0) = a_j\delta D_0\alpha^p$$

(since $D_i\alpha^p = 0$ if $i \neq 0$). It follows that it will suffice to show that $a_j\delta D_0\alpha^p$ is in $\mathfrak{O}$ for each j as this will demonstrate the regularity of each ω_j at $\mathfrak{p}$.

If $K' = K(\beta)$, then we define a mapping S_β of K' into K as follows: if $a \in K'$ and $a = a_0 + a_1\beta + \cdots + a_{p-1}\beta^{p-1}$ with $a_i \in K$, then $S_\beta(a) = a_{p-1}$ (compare with [3]). We observe that $a = \sum_{\nu=0}^{p-1} S_\beta(a\beta^{p-1-\nu})\beta^\nu$ for all $a \in K'$; and furthermore, if β, γ are two generators of K' over K with D'_β, D'_γ denoting the derivations of K' with respect to β, γ respectively then, ([3]), we have $S_\beta(a) = S_\gamma(a(D'_\gamma\beta)^{1-p}) = S_\gamma(a(D'_\beta\gamma)^{p-1})$ for all a in K'. We let $S = S_\alpha$ and $T = S_t$. Then:

$$\sum_{j=0}^{p-1} a_j\delta(D'_t\alpha)\alpha^j = \sum_{j,\nu} \{a_j\delta T((D'_t\alpha)t^{p-1-\nu}\alpha^j)\}t^\nu \in \mathfrak{O}' \tag{1}$$

and therefore, since $\mathfrak{O}' = \mathfrak{O} + \mathfrak{O}t + \cdots + \mathfrak{O}t^{p-1}$, for $\nu = 0, \cdots, p-1$,

$$\sum_{j=0}^{p-1} a_j\delta T((D'_t\alpha)t^{p-1-\nu}\alpha^j) = \sum_{j=0}^{p-1} a_j\delta S((D'_t\alpha)^p t^{p-1-\nu}\alpha^j) \in \mathfrak{O}.$$

It follows that, for $\nu = p-1$, $\sum_{j=0}^{p-1} a_j\delta S((D'_t\alpha)^p\alpha^j)$ is in $\mathfrak{O}$. However, $S((D'_t\alpha)^p\alpha^j) = 0$ for $j = 0, \cdots, p-2$ and $S((D'_t\alpha)^p\alpha^{p-1}) = (D'_t\alpha)^p$. Thus, we have $a_{p-1}\delta(D'_t\alpha)^p$ is in $\mathfrak{O}$. If we multiply the left-hand side of equation (1) by $\alpha^{p-1-\mu}$, where $\mu \in \{0, \cdots, p-1\}$, we get:

$$\sum_{j=0}^{p-1} a_j\delta(D'_t\alpha)\alpha^{j+p-1-\mu} = \sum_{i=0}^{p-1-\mu-1} a_{\mu+1+i}\delta\alpha^p(D'_t\alpha)\alpha^i + \sum_{i=p-1-\mu}^{p-1} a_{\mu+i+1-p}\delta(D'_t\alpha)\alpha^i \tag{2}$$

is in $\mathfrak{O}'$. Applying the argument, which we used to show that $a_{p-1}\delta(D'_t\alpha)^p$

was in $\mathfrak{D}$, to the right hand side of equation (2), we see that $a_{\mu+i+1-p}\delta(D'_t\alpha)^p$ is in $\mathfrak{D}$ for $\mu = 0, \cdots, p-1$. Hence it will suffice to show that $v_{\mathfrak{p}}(a_j\delta(D'_t\alpha)^p) = v_{\mathfrak{p}}(a_j\delta(D_0\alpha^p))$ i.e., that $v_{\mathfrak{p}}((D'_t\alpha)^p) = v_{\mathfrak{p}}((D_0\alpha^p))$.

Now $\alpha = y_0 + y_1t + \cdots + y_{p-1}t^{p-1}$ with the y_i in $\mathfrak{D}$ and thus, we have:

$$\begin{aligned}
\alpha^p &= y_0{}^p + y_1{}^p\eta_0 + \cdots + y_{p-1}{}^p\eta_0{}^{p-1}\\
(D'_t\alpha) &= y_1 + 2y_2t + \cdots + (p-1)y_{p-1}t^{p-2}\\
(D'_t\alpha)^p &= y_1{}^p + 2^py_2{}^p\eta_0 + \cdots + (p-1)^py_{p-1}{}^p\eta_0{}^{p-2}\\
(D'_0\alpha^p) &= y_1{}^p + 2y_2{}^p\eta_0 + \cdots + (p-1)y_{p-1}{}^p\eta_0{}^{p-2}.
\end{aligned}$$

Since $v_{\mathfrak{p}}(y_i) \geqq 0$ for $i = 0, \cdots, p-1$, we also see that

$$v_{\mathfrak{p}}((D'_t\alpha)^p) = \min\{pv_{\mathfrak{p}}(y_1), pv_{\mathfrak{p}}(y_2) + 1, \cdots, pv_{\mathfrak{p}}(y_{p-1}) + p - 2\} = v_{\mathfrak{p}}((D_0\alpha^p)).$$

The proof of our proposition is now complete.

We remark that, in the course of our proofs of Propositions 3, 4 and 5, we have established the following lemma which will be used quite frequently in the next section.

LEMMA 1. *Let K, k, K', k' be as in Theorem 1; let $\mathfrak{p}$ be a prime divisor of K/k; let $\mathfrak{p}'$ be a prime divisor of K'/k' which lies over $\mathfrak{p}$; and let ω be a differential of K/k. If the differential* $\operatorname{cotr}\omega$ *of K'/k' is regular at $\mathfrak{p}'$, then ω is regular at $\mathfrak{p}$.*

3. The main result for arbitrary extensions.

Let k' be an arbitrary extension of k (which is linearly disjoint from K over k). The main object of this section is to extend our Theorem 1 to this situation. In other words, we wish to prove the following theorem:

THEOREM 2. *If K/k is a regular extension, if k' is an extension of k which is linearly disjoint from K/k, and if K'/k' is obtained from K/k by extending the ground field from k to k', then $p_g(K'/k') \leqq p_g(K/k)$ with equality if k' is a separable extension of k.*

Proof. Let $\{a_\alpha\}$ be a vector basis of k' over k and let $\bar{\omega}$ be an element of $\Omega(K'/k')$. Then there exist $a_{\alpha_1}, \cdots, a_{\alpha_n}$ (where, to be sure, n depends on $\bar{\omega}$) such that $\bar{\omega} = \sum^{n} a_{\alpha_i} \operatorname{cotr}\omega_i$ with the ω_i differentials of K over k. As in the previous section, it will suffice to show that $\omega_i \in \Omega(K/k)$ for all i in order to prove $p_g(K'/k') \leqq p_g(K/k)$. Let $\bar{k} = k(a_{\alpha_1}, \cdots, a_{\alpha_n})$. Then, if $\bar{K}/\bar{k}$ is obtained from K/k by extending the ground field to $\bar{k}$, we observe that $\bar{\omega}$ may be viewed as a differential of $\bar{K}/\bar{k}$ and that it will suffice to show that $\bar{\omega}$ (viewed in this way) is in $\Omega(\bar{K}/\bar{k})$ since the result will then follow from our proof of Theorem 1. In other words, we may assume that $\bar{\omega} = \operatorname{cotr}\omega$. We will consider two cases, namely: (a) k' is separable over k (b) k' is purely inseparable over k. It is clearly sufficient to do so.

Case (a). We shall first establish that $p_g(K/k) \leqq p_g(K'/k')$ in this case so that the equality assertion will follow from $p_g(K'/k') \leqq p_g(K/k)$. To do this, we let $\omega'_1 = \operatorname{cotr} \omega_1$ where $\omega_1 \in \Omega(K/k)$ and we let $\mathfrak{p}'$ be a prime divisor of K'/k'. If $\mathfrak{p}'$ is trivial on K, then the regularity of ω'_1 at $\mathfrak{p}'$ is an immediate consequence of Remark 6. If $\mathfrak{p}'$ is not trivial on K, we let $\mathfrak{p}$ be the contraction of $\mathfrak{p}'$ to K. Then there exist p-parameters $u_1, \cdots, u_\rho$ of p and elements $u_{\rho+1}, \cdots, u_\sigma$ of k' such that $u_1, \cdots, u_\sigma$ are p-parameters of $\mathfrak{p}'$. To see this, we observe that p-independent elements of k (over k^p) remain p-independent over k'^p in view of our separability assumption. So if $\xi_1, \cdots, \xi_{\rho+r}$ are uniformizing coordinates of $\mathfrak{p}$ and if $\omega_1 du_1 \cdots du_\rho = A d\xi_1 \cdots d\xi_{\rho+r}$ with $A \in \mathfrak{O}_\mathfrak{p}$, then $\operatorname{cotr} \omega_1 du_1 \cdots du_\rho du_{\rho+1} \cdots du_\sigma = A d\xi_1 \cdots d\xi_{\rho+r} du_{\rho+1} \cdots du_\sigma$ and our assertion follows from Remark 6.

We now wish to show that $p_g(K/k) \geqq p_g(K'/k')$ in this case. Let $\mathfrak{p}$ be a prime divisor of K/k and let $\mathfrak{p}'$ be a prime divisor of K'/k' which lies over $\mathfrak{p}$. Then we can choose p-parameters $u_1, \cdots, u_\sigma$ of $\mathfrak{p}'$ such that $u_1, \cdots, u_\rho$ are p-parameters of $\mathfrak{p}$ (compare with previous paragraph). Then if $\xi_1, \cdots, \xi_{\rho+r}$ are uniformizing coordinates at $\mathfrak{p}$, we have $\omega du_1 \cdots du_\rho = A d\xi_1 \cdots d\xi_{\rho+r}$ and $(\operatorname{cotr} \omega) du_1 \cdots du_\sigma = A d\xi_1 \cdots d\xi_{\rho+r} du_{\rho+1} \cdots du_\sigma$. Suppose for the moment, that $\sigma = \rho$; i.e., that $u_1, \cdots, u_\rho$ are p-parameters of $\mathfrak{p}'$, then A will be in $\mathfrak{O}_{\mathfrak{p}'} \cap K = \mathfrak{O}_\mathfrak{p}$ since we have $\mathfrak{O}_{\mathfrak{p}'}$ as a quotient ring of $k'[\mathfrak{O}_\mathfrak{p}]$ and the derivations D_i of K such that $D_i \xi_j = \delta_{ij}$ can be extended to derivations of K' which map $k'[\mathfrak{O}_\mathfrak{p}]$ into $k'[\mathfrak{O}_\mathfrak{p}]$ (i.e., the ξ_i are uniformizing coordinates of $\mathfrak{p}'$). Now, if $\sigma > \rho$, then we adjoin $u_{\rho+1}, \cdots, u_\sigma$ to k and we let $\mathfrak{p}^*$ be the contraction of $\mathfrak{p}'$ to $K(u_{\rho+1}, \cdots, u_\sigma) = K^*$. It follows that if ω^* is the cotrace of ω from K to K^*, then $\bar{\omega}$ is the cotrace of ω^* from K^* to K' and, moreover, ω^* is regular at $\mathfrak{p}^*$. We let $k^* = k(u_{\rho+1}, \cdots, u_\sigma)$ and we observe that K^*/k^* is obtained from K/k by extending the ground field from k to k^*. It follows from our Lemma 1 that ω is regular at $\mathfrak{p}$. The proof of our theorem is now complete in this case.

Case (b). Let $\mathfrak{p}$ be a prime divisor of K/k and let $\mathfrak{p}'$ be the prime divisor of K'/k' which lies over $\mathfrak{p}$. Let η_1 be a uniformizing parameter of $\mathfrak{p}'$. We observe that there exist p-parameters $v_1, \cdots, v_\sigma$ of $\mathfrak{p}'$ such that if K^* is obtained from K by extending the ground field to $k^* = k(v_1, \cdots, v_\sigma)$, then $\eta_1 \in K^*$.[6] Let $\mathfrak{p}^*$ be the contraction of $\mathfrak{p}'$ to K^* and let $v_1, \cdots, v_\sigma, u_1, \cdots, u_\rho$

---

[6] To see this, we let $w_1, \cdots, w_\mu$ be p-parameters of $\mathfrak{p}'$ with k'_1 as a field of p-constants. Then $\eta_1 \in k'_1[w_1, \cdots, w_\mu, K]$ i. e, $\eta_1 = f(w_1, \cdots, w_\mu, z_1, \cdots, z_\mu)$ where $z_i \in K$ and f is a polynomial with coefficients in k'_1. We set $k^* = k(w_1, \cdots, w_\mu,$ coefficients of $f)$.

be p-parameters of $\mathfrak{p}^*$ where we assume that the u_i are all in k'^p.[7] It follows that if $k^*_1 = k \cap k'_1$ where k'_1 is a field of p-constants for the v_i, then there exists a field of p-constants for $\{v_i, u_j\}$, k^*_2, such that $k^*_2 \subset k^*_1$ (e.g., take k^*_2 as a maximal subfield of k^*_1 containing k^p such that $[k^*_2(u_1, \cdots, u_\rho) : k^*_2] = p^\rho$). Let ω^* denote the cotrace of ω taken from K to K^* and let $\eta_1, \cdots, \eta_{\sigma+r}$ be uniformizing coordinates of $\mathfrak{p}'$ relative to the v_i and k'_1 such that the η_i are in K^* (compare with the proof of [1] Proposition 1). Then we observe that K^* is separable algebraic over $k^*_1(\eta_1, \cdots, \eta_{\sigma+r})$ since K' is separable algebraic over $k'_1(\eta_1, \cdots, \eta_{\sigma+r})$, k'_1 is purely inseparable over k^*_1, k'_1 and K^* are linearly disjoint over k^*_1, and $K' = k'_1(\eta_1, \cdots, \eta_{\sigma+r}, y_1, \cdots, y_N)$ where $K^* = k^*_1(\eta_1, \cdots, \eta_{\sigma+r}, y_1, \cdots, y_N)$ ([2a], Corollary 41, p. 125 is used here). It follows that we have derivations $D^*_1, \cdots, D^*_{\sigma+r}$ of K^*/k^*_1 such that $D^*_i\eta_j = \delta_{ij}$ $(i, j = 1, \cdots, \sigma + r)$ and also that these derivations can be extended to derivations $D'_1, \cdots, D_{\sigma+r}$ of K'/k'_1. Consequently, noting that the cotrace of ω^* taken from K to K' is $\tilde{\omega}$, we have: $\tilde{\omega} dv_1 \cdots dv_\sigma = A\, d\eta_1, \cdots, d\eta_{\sigma+r}$ and $\omega^* dv_1 \cdots dv_\sigma du_1 \cdots du_\rho = A\, d\eta_1 \cdots d\eta_{\sigma+r} du_1 \cdots du_\rho$ with A in K^*. However, $\tilde{\omega}$ is regular at $\mathfrak{p}'$ and therefore, A is in $\mathfrak{O}_{\mathfrak{p}'} \cap K^* = \mathfrak{O}_{\mathfrak{p}^*}$. It follows that ω^* is regular at $\mathfrak{p}^*$ and consequently, that ω is regular at $\mathfrak{p}$ (Lemma 1). The proof is now complete.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] O. Zariski and P. Falb, "On differentials in function fields," *American Journal of Mathematics*, vol. 83 (1961), pp. 542-556.

[2a] O. Zariski and P. Samuel, *Commutative algebra*, Vol. I; D. Van Nostrand Co., Princeton, 1958.

[2b] O. Zariski and P. Samuel, *Commutative algebra*, Vol. II; D. Van Nostrand Co., Princeton, 1960.

[3] J. Tate, "Genus change in inseparable extensions of function fields," *Proceedings of the American Mathematical Society*, vol. 3 1952), pp. 400-406.

---

[7] We can make this assumption for the reason that if $u_1, \cdots, u_T$ (say) are not in k'^p, then a p-basis, $\{w_1, \cdots, w_\mu\}$, of $k'^p(u_1, \cdots, u_T, v_1, \cdots, v_\sigma)$ will be p-parameters of $\mathfrak{p}'$ and $u_{T+1}, \cdots, u_\rho, w_1, \cdots, w_\mu$ will be p-parameters of $\mathfrak{p}^*$.

FACTORS OF CUBES.*

By Kyung Whan Kwun and Frank Raymond.

Introduction. Let E^n denote euclidean n-space, I^n denote the n-cube or n-cell, and S^n denote the n-sphere, or what is the same, the boundary of I^{n+1}. We shall use the equality sign between two topological spaces to mean topological equivalence. Consider the equation $I^n = A \times B$. If $n \leqq 4$, then both A and B must be cells [17] and [4]. On the other hand, if $n \geqq 5$, examples of $(n-1)$-combinatorial manifolds M with boundary such that $M \times I = I^n$ and $M \neq I^{n-1}$ have been given by Poenaru [12], Mazur [11], and Curtis [8]. For $n \geqq 6$ the argument simply consists in finding a contractible combinatorial $(n-1)$-manifold M with boundary whose boundary is not simply connected. Since the boundary of $M \times I$ must be simply connected, it follows from Smale's proof of the generalized Poincaré conjecture that $M \times I = I^n$. No other unorthodox factorization of cubes were known heretofore (at least to our knowledge). In fact, it is not difficult to see that if M^{n-1} is a combinatorial $(n-1)$-manifold with boundary such that $M^{n-1} \times I^2 = I^{n+1}$, then $M^{n-1} \times I = I^n$, for all n except perhaps 5 or 6, see [8]. Thus it seemed to be an interesting problem to construct, if possible, factorizations of the cube where the factors fail to be manifolds with boundary. By using a recent result of Andrews and Curtis, we construct (Theorem 2) for each $n \geqq 3$, an n-dimensional space X such that $X \times I$ is not a manifold and $X \times I^2 = I^{n+2}$. With this result it is not difficult to factor $I^n = A \times B$ so that *each* factor fails to be a manifold with boundary if $n \geqq 8$. The spaces X that we construct in Theorem 2 are modified in Theorem 3 so that for each $n \geqq 4$ we have n-dimensional spaces M' such that M' fails to be a manifold with boundary at exactly one point and $M' \times I = I^{n+1}$. This enables us to construct explicit 4 and 5-dimensional factors of 5 and 6-cubes which are *irreducible* in the sense that they can not be further factored.

Fortunately, there is a limit to the pathology exhibited by factors of cells. We show in Theorem 1 that any factor of a cube must be indistinguishable from a cell on the homology level both locally and globally. That is, we show that a generalized-cell can be written as a product if and only if each factor is a generalized cell. (Every cell is, of course, a generalized cell and also a

---

* Received January 4, 1962.

generalized cell of dimension 2 or less is an actual cell.) The significance of this result for our constructions is to enable us to attach a meaning to "boundary" and "interior" of factors of cubes and candidates for factors of cubes.

1. Factorization of generalized cells. In this section we shall use Čech cohomology and Borel-Moore homology theory with integer coefficients and *compact supports.* For the definitions, as we plan to use them, of generalized manifolds (cohomology manifolds) or generalized manifolds with boundary the reader is referred to [13].

1.1. *Definition.* A *generalized n-cell X* will mean a compact connected orientable generalized n-manifold X with non-empty boundary such that $H^p(X) = 0$, $p > 0$.

The reader is referred to [14; 7] where the Poincaré duality is used to obtain the Lefschetz duality theorem for generalized manifolds with boundary. The point is that Poincaré duality implies that $H_*(X, X - \operatorname{Bd} X) = 0$, where $\operatorname{Bd} X$ denotes the boundary of X. The universal coefficient theorem implies $H_p(X) = 0$, $p > 0$. Therefore, by the preceding remark $H_p(X - \operatorname{Bd} X) = 0$, $p > 0$ and hence by Poincaré duality $H^p(X - \operatorname{Bd} X) = 0$, $p \neq n$. The exact cohomology sequence now implies that $H^p(\operatorname{Bd} X) \approx H^p(S^{n-1})$, where S^{n-1} denotes the $(n-1)$-sphere. Hence,

Every generalized n-cell is an acyclic generalized n-manifold with sphere-like boundary. In particular, a separable metric generalized n-cell is an actual euclidean n-cell if $n \leqq 2$. *A generalized 3-cell X must have a 2-sphere boundary although the boundary does not necessarily fit onto X as a locally euclidean manifold with boundary.*

We point out here if we use coefficient from a *field* instead of the integers then the discussion shows that this modified definition of generalized cell agrees with that originally given by Wilder [15; p. 287]. In fact, for our applications of Theorem 1 it would be sufficient to only consider coefficients from a field.

THEOREM 1. *Let* $X = A \times B$. *Then X is a generalized n-cell if and only if A is a generalized k-cell and B is a generalized* $(n-k)$*-cell.*

Proof. The Künneth theorem immediately implies that X is acyclic if and only if both A and B are acyclic. By [13; Theorem 6], X is a generalized

n-manifold with boundary if and only if A and B are generalized k and $(n-k)$ manifolds with possible boundary. Since both $\operatorname{Int} A$ and $\operatorname{Int} B$ ($\operatorname{Int} X$, means the interior of X or $X-\operatorname{Bd} X$) must be orientable, A and B can not be acyclic without having a non-empty boundary. Thus by the previous remarks X, A and B are generalized cells.

Recall that for a generalized manifold X with boundary, $\operatorname{Bd} X$ is the set of points for which the n-th local co-betti number vanishes and that $\operatorname{Int} X = X - \operatorname{Bd} X$. We remark that in [13; Theorem 6] it is shown that $\operatorname{Bd}(A \times B) = (\operatorname{Bd} A \times B) \cup (A \times \operatorname{Bd} B)$ where $(\operatorname{Bd} A \times B) \cap (A \times \operatorname{Bd} B) = \operatorname{Bd} A \times \operatorname{Bd} B$. Furthermore, $\operatorname{Bd} A \times \operatorname{Bd} B$ (a generalized $(n-2)$-manifold) fits onto both $\operatorname{Bd} A \times B$ and $A \times \operatorname{Bd} B$ as a manifold with boundary; $\operatorname{Int}(A \times B)$ is precisely $\operatorname{Int} A \times \operatorname{Int} B$.

As a corollary of Theorem 1 we obtain the results of Young [17] and Bing [4].

COROLLARY. *Let $A \times B = I^n$. Then A and B are cells, if $n \leqq 4$.*

Proof. If $n \leqq 3$, or if $n = 4$ and one factor is 2-dimensional, then we need only assume $A \times B$ is a generalized n-cell. In this situation, the corollary is a direct consequence of Theorem 1 and the fact that low dimensional generalized cells are actual cells. In the remaining case, we may assume, by the theorem, that B is 1-dimensional and therefore a 1-cell,

$$\operatorname{Bd}(A \times I) = (\operatorname{Bd} A \times I) \cup (A \times I) = (S^2 \times I) \cup (A \times 0) \cup (A \times I) = S^3.$$

Since $(A \times 0) \subset S^3$ and $\operatorname{Bd} A = S^2$, A is a 3-cell by Bing's theorem [3] if we can show that $\operatorname{Int}(A \times 0)$ is uniformly locally simply connected. The uniform local contractibility of $\operatorname{Int} A$ is implied by that of $\operatorname{Int} I^4$.

2. Factors of cells. In [1], using the technique of Bing [2], Andrews and Curtis show that if α is an arc in E^n and if α is collapsed to a point then the resulting space when multiplied by E^1 is the end result of a certain pseudo isotopy of E^{n+1} onto itself. The result of Andrews and Curtis, when specialized in the cube, yields:

LEMMA 1. *Let α be an arc in $\operatorname{Int} I^n$, U a neighborhood of α, and $\epsilon > 0$. Then there exists a pseudoisotopy μ_t of $I^n \times E^1$ onto itself such that*

(1) *μ_0 is the identity,*

(2) *μ_t, for each $t < 1$, is a homeomorphism,*

(3) *μ_t, $0 \leqq t \leqq 1$, is the identity on $(I^n - U) \times E^1$,*

(4) *$\{\mu_1^{-1}(p) \mid p \in I^n \times E^1\} = \{A \times y \mid A = \alpha$ or a point of $I^n - \alpha$ and $y \in E^1\}$,*

(5) *During the isotopy no point of* $I^n \times E^1$ *is moved along the* E^1 *direction by as much as* ϵ.

LEMMA 2. *Exactly the same as Lemma* 1 *except that* S^1 *replaces* E^1.

Proof. We merely replace the vertical subdivision of E^1 by a cyclic subdivision of S^1 (cf. [2], p. 410, the first paragraph).

We are now ready to prove the main result.

THEOREM 2. *Let* X *be the space obtained from* I^n *by shrinking an arc* α, $\alpha \subset \operatorname{Int} I^n$, *to a point. Then* $X \times I^2 = I^{n+2}$.

Proof. We show that (1) $X \times I^2$ is a manifold with boundary, (2) $\operatorname{Bd}(X \times I^2) = S^{n+1}$, and (3) $X \times I^2 \subset E^{n+2}$. Then it follows that $X \times I^2$ is the closure of a bounded complementary domain in E^{n+2} of S^{n+1}. By Brown [7], S^{n+1} is collared in $X \times I$, and by Brown [6], $X \times I^2$ is an $(n+2)$-cell.

First, X is a generalized n-cell with an actual $(n-1)$-sphere as boundary. That X is a generalized n-cell over every field follows from Wilder's monotone mapping theorem, [16], and hence by a theorem of the authors [10], X is a generalized cell over the integers. Thus, $X \times I^2$ is a generalized $(n+2)$-cell and $\operatorname{Int} X$ and $\operatorname{Bd} X$ have a definite meaning.

(1) We show that each point of $X \times I^2$ has a closed neighborhood homeomorphic to an $(n+2)$-cell. Now, $\operatorname{Int}(X \times I^2) = \operatorname{Int} X \times \operatorname{Int} I^2 = \operatorname{Int} X \times E^2 = (\operatorname{Int} X \times I^1) \times E^1 = E^{n+2}$ by Andrews and Curtis [1]. Let $p \in \operatorname{Bd}(X \times I^2)$. If $p \in \operatorname{Bd} X \times I^2$, p clearly has the desired neighborhood. Suppose $p \in X \times \operatorname{Bd} I^2$. Then p has a closed neighborhood homeomorphic to $(X \times S^1) \times I = I^{n+1} \times S^1$ which itself is a manifold with boundary.

(2) Let G be the upper semi-continuous decomposition of I^n consisting of α and the points of $I^n - \alpha$. Let ϕ denote the quotient map of G. Let μ_t be a pseudo-isotopy of $I^n \times S^1$ onto itself such that μ_t is the identity on $\operatorname{Bd} I^n \times S^1$ and $\{\mu_1^{-1}(x) \mid x \in I^n \times S^1\} = G \times S^1$. The existence of μ_t is assured by Lemma 2. Consider the maps $f\colon \operatorname{Bd}(I^n \times I^2) \to Bd(I^n \times I^2)$ and $g\colon \operatorname{Bd}(I^n \times I^2) \to \operatorname{Bd}(X \times I^2)$ defined as follows:

$$f(x,y) = \begin{cases} (x,y), & \text{if } x \in \operatorname{Bd} I^n, y \in I^2 \\ \mu_1(x,y), & \text{if } X \in I^n, y \in \operatorname{Bd} I^2 \end{cases}$$

$$g(x,y) = (\phi(x), y).$$

The maps f and g are continuous and onto. Furthermore their point inverses are precisely $\{A \times y \mid A \in G, y \in I^2, A \times y \subset \operatorname{Bd}(I^n \times I^2)\}$. Since $\operatorname{Bd}(I^n \times I^2)$

and $\mathrm{Bd}(X \times I^3)$ are the decomposition spaces of the same upper semicontinuous decomposition, they are homeomorphic.

(3) $X \times I^2 \subset X \times E^2 \subset E^{n+2}$.

COROLLARY. *For each* $n \geqq 4$, *there exists a space* M^n *which is not a manifold with boundary and such that* $M^n \times I = I^{n+1}$.

Proof. Let X^{n-1} of the theorem be obtained by using an arc α of Blankinship [5]. Then X^{n-1} fails to be locally euclidean at $\phi(\alpha)$. Hence, $X^{n-1} \times I = M^n$ is not a manifold with boundary as $\mathrm{Bd}\, M^n$ fails to be locally euclidean at two points. But $M^n \times I = I^{n+1}$ by the theorem.

Remarks. By modifying the method of [9], we can also construct a generalized $(n-1)$-cell X such that X contains *no* $(n-1)$-cell but $X \times I^2 = I^{n+1}$, for all $n \geqq 4$. Many questions concerning factors of cubes can be raised. The following seem interesting.

Question 1. Does there exist a space M such that $M \times I = I^{n+1}$ and $\mathrm{Bd}\, M$ contains no $(n-1)$-cell?

Question 2. Does $M \times I = I^{n+1}$ imply $M \times M = I^{2n}$? If X is the space in Theorem 2, is $X \times X$ a cube?

Having obtained the Corollary one can make many remarks about pathological involutions on I^n and S^n. In particular, one can obtain, for all $n \geqq 5$, involutions on I^n such that the orbit space is still I^n and the fixed point set fails to be a manifold with boundary.

Let us consider possible factors of I^n. Young and Bing (the Corollary to Theorem 1), have shown that the factors are necessarily cells if $n \leqq 4$. In particular, a 3-dimensional factor is unique. But Theorem 2 shows that this is not the case if $n \geqq 5$. The method of Theorem 2 and its Corollary easily implies in case $n \geqq 8$, $I^n = A \times B$, where neither A nor B is a manifold with boundary. We conjecture that this is also possible when $n = 6$ or 7.

3. Modifications of *M*. Let $A \times B = C$. We call A an *irreducible* factor of C if A does not permit a non-trivial factorization. Since our M is not irreducible, one might ask if each $(n-1)$-dimensional irreducible factor of I^n is necessarily a manifold with boundary. The following construction of M' not only simplifies M in the sense that M' has only one point at which it fails to be a manifold with boundary (instead of two) but also shows that I^n, $n = 5$, 6 can have "bad" irreducible factors. This problem was posed by Bing in the course of conversation.

Construction of M'. Let $M = X \times I$ as before. Let $I_1{}^n = [\frac{1}{4}, \frac{3}{4}]^n$ and assume $\alpha \subset \operatorname{Int} I_1{}^n$. Let $X' = \phi(I_1{}^n)$. We consider S^1 as the real numbers modulo *three*. Let μ_t be the pseudo isotopy of Lemma 2, with $I_1{}^n$ replacing I^n and with $\epsilon = \frac{1}{2}$. Consider the continuous map h of $I^n \times S^1$ onto itself defined by

$$h(x, y) = \begin{cases} (x, y) & \text{if } x \in I^n - \operatorname{Int} I_1{}^n \\ \mu_1(x, y) & \text{if } x \in I_1{}^n. \end{cases}$$

Let k be a homeomorphism of $I^n \times S^1$ onto $X \times S^1$ defined as the composite

$$\begin{tikzcd} (x, y) \arrow[r, "h^{-1}"] & (A, y') \arrow[r, "(\phi \times \text{identity})"] & (\phi(A), y'). \end{tikzcd}$$

Observe that h moves the S' coordinate by less than $\frac{1}{2}$ and if $x \in \operatorname{Bd} I_1{}^n$, then $y' = y$. Let M' be the subset of $X \times S^1$ bounded by

$$\operatorname{Bd} M' = (\operatorname{Bd} X \times I) \cup (X \times 0) \cup k(I^n \times 1).$$

Clearly, M' has only one "bad" point, viz. $\phi(\alpha) \times 0$.

THEOREM 3. $M' \times I = I^{n+2}$.

Proof. (1) $M' \times I$ is a manifold with boundary. This is so because each point of M' is locally like a point of M.

(2) $\operatorname{Bd}(M' \times I) = S^{n+1}$. Consider the $(n+2)$-cell $Y = (X \times [0, 2]) \times I$. Observe that $\operatorname{Bd}(k(I^n \times 1) \times I)$ is locally flatly imbedded in $\operatorname{Bd} Y = S^{n+1}$, and that $\operatorname{Bd}(M' \times I)$ is the union of one of the complementary domains of $\operatorname{Bd}(k(I^n \times 1) \times I)$ in $\operatorname{Bd} Y$ with $k(I^n \times I) \times I$. Since both are $(n+1)$-cells, $\operatorname{Bd}(M' \times I) = S^{n+1}$ (See Figure 1).

(3) $M' \times I \subset E^{n+2}$. This follows from the fact that

$$M' \times I \subset (X \times [0, 2] \times I = I^{n+2}.$$

COROLLARY. *For I^n, $n = 5, 6$, there exists an irreducible $(n-1)$-dimensional factor which is not a manifold with boundary.*

Proof. Let $I^n = M' \times I$, with the arc α being an arc of Blankinship. Suppose $M' = A \times I^k$, for some $k > 0$. Then, using a certain homogeneity of the cell, it can be shown that M' cannot have only *one* "bad" point. Thus, for all n, M' is irreducible with respect to an arc. In particular, if $n = 5$ or 6, any non trivial factor of M' would have to be an actual cell by Theorem 1.

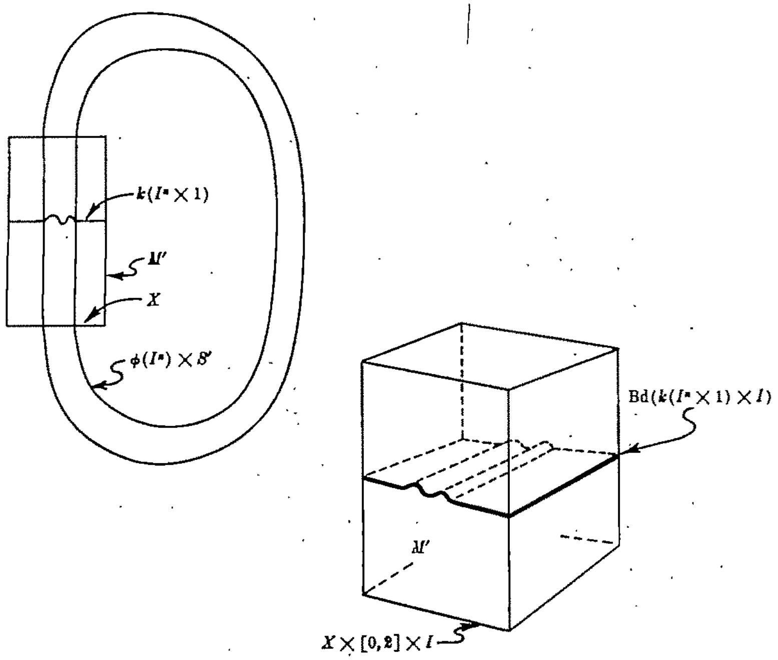
$k(I^n \times 1)$
M'
X
$\phi(I^n) \times S'$
$\mathrm{Bd}(k(I^n \times 1) \times I)$
M'
$X \times [0,2] \times I$

FIGURE 1.

SEOUL NATIONAL UNIVERSITY AND
UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] J. J. Andrews and M. L. Curtis, "N-space modulo an arc," to appear in *Annals of Mathematics.*

[2] R. H. Bing, "The cartesian product of a certain non-manifold and a line is E^4," *Annals of Mathematics*, vol. 70 (1959), pp. 399-412.

[3] ———, "A surface is tame if its complement is 1-ULC," *Transactions of the American Mathematical Society*, vol. 101 (1961), pp. 294-305.

[4] ———, "A set is a 3-cell if its cartesian product with an arc is a 4-cell," *Proceedings of the American Mathematical Society*, vol. 12 (1961), pp. 13-19.

[5] W. Blankinship, "Generalization of a construction of Antoine," *Annals of Mathematics*, vol. 53 (1951), pp. 276-297.

[6] Morton Brown, "A proof of the generalized Schönflies Theorem," *Bulletin of the American Mathematical Society*, vol. 66 (1960), pp. 74-76.

[7] ———, "Locally flat embeddings of topological manifolds," *to appear.*

[8] M. L. Curtis, "Cartesian products with intervals," *Proceedings of the American Mathematical Society*, vol. 12 (1961), pp. 819-820.

[9] Kyung Whan Kwun, "Factors of E^n," to appear in *Michigan Mathematical Journal.*

[10] Kyung Whan Kwun and Frank Raymond, "Generalized cells in generalized manifolds," *Proceedings of the American Mathematical Society*, vol. 11 (1960), pp. 135-139.

[11] Barry Mazur, "A note on some contractible 4-manifolds," *Annals of Mathematics*, vol. 73 (1961), pp. 221-228.

[12] V. Poenaru, "La decomposition de l'hypercube en produit topologique," *Bulletin de la Société Mathematique de France*, vol. 88 (1960), pp. 485-488.

[13] Frank Raymond, "Union and separation theorems for generalized manifolds with boundary," *Michigan Mathematical Journal*, vol. 7 (1960), pp. 7-21.

[14] ———, "Local cohomology group with closed supports," *Mathematische Zeitschrift*, vol. 76 (1961), pp. 31-41.

[15] R. L. Wilder, *Topology of manifolds*, American Mathematical Society Colloquium publication 32 (1949).

[16] ———, "Monotone mappings of manifolds II," *Michigan Mathematical Journal*, vol. 5 (1958), pp. 19-23.

[17] G. S. Young, "On the factors and fiberings of manifolds," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 215-223.

HERMITIAN FORMS OVER LOCAL FIELDS.*

By Ronald Jacobowitz.

One of the basic problems in the theory of forms is the determination of invariants for equivalence of two forms in an arbitrary field. Let F be any field, not necessarily commutative, with an anti-automorphic involution, i.e., a bijection $a \to a^*$ of F into itself, satisfying $(a+b)^* = a^* + b^*$, $(ab)^* = b^*a^*$, and $a^{**} = a$ for all a, b in F. Then, in analogy with the complex numbers, we may call an $n \times n$ matrix $A = (a_{\lambda\mu})$ over F *conjugate-symmetric* if $a_{\lambda\mu} = a_{\mu\lambda}^*$ for all λ, μ, and we define the *n-ary hermitian form* (in the indeterminates $\xi_1, \cdots, \xi_n$) corresponding to A to be the expression $h(\xi_1, \cdots, \xi_n) = \sum_\lambda \sum_\mu \xi_\lambda a_{\lambda\mu} \xi_\mu^*$. (Of course, if $a^* = a$ for all a, A is a *symmetric* matrix and h becomes a *quadratic* form.)

Given an arbitrary $n \times n$ matrix $R = (r_{\lambda\mu})$ over F, the matrix $B = RA\,{}^tR^*$ (${}^tR^* =$ conjugate-transpose of R) is also conjugate symmetric; thus if we make the linear substitution $\xi_\lambda = \sum_\nu \eta_\nu r_{\nu\lambda}$ $(1 \leqq \lambda \leqq n)$ into $h(\xi_1, \cdots, \xi_n)$, the $\langle \eta_\nu \rangle$ indeterminates, there results the n-ary herimitian form $h'(\eta_1, \cdots, \eta_n) = \sum_\nu \sum_\rho \eta_\nu b_{\nu\rho} \eta_\rho^*$, with coefficients $b_{\nu\rho} = \sum_\lambda \sum_\mu r_{\nu\lambda} a_{\lambda\mu} r_{\rho\mu}^*$. The forms h and h' are called *(fractionally) equivalent* if there exists an *invertible* matrix $R = (r_{\lambda\mu})$ over F for which the above substitution takes h into h', i.e., if $(b_{\lambda\mu}) = RA\,{}^tR^*$. Furthermore, if there is a reasonable definition of a ring $\mathfrak{O}$ of *integers* in F, and if the matrix R is *unimodular* (i.e., the entries of both R and R^{-1} lie in $\mathfrak{O}$), then h and h' will be called *integrally equivalent.*

In this paper, we shall obtain invariants for integral equivalence over commutative and certain non-commutative local fields of characteristic not 2; the fractional invariants were found by Jacobson [6], and will be described in § 3. The similar problem for integral equivalence of *quadratic* forms over a commutative local field has been completely solved by Durfee [5] and O'Meara [8], [9].

Considerable use will be made of many of the concepts and proofs found in O'Meara's papers and forthcoming book [10] and, indeed, we feel greatly indebted to O'Meara for his generous encouragement and assistance. We wish also to express thanks to C.-H. Sah for several stimulating conversations.

---

* Received November 30, 1961.

1. Local fields with an involution. Throughout this paper, F will denote a field of characteristic not 2, with the following properties:

1) F has an involution $a \rightarrow a^*$ different from the identity.

2) F is *local*, i. e., is complete and has finite residue class field under a discrete non-archimedean valuation, written $|a|$. (for discussion of these ideas, cf. Ch. I-III of [1]).

3) If F is non-commutative, we require that F be a quaternion algebra over a commutative subfield E. As shown in [4], p. 135, F is a 4-dimensional vector space over E, with basis $\langle 1, u, v, uv \rangle$ satisfying $u^2 = \theta$, $v^2 = \Delta$, $uv = -vu$, where θ and Δ are any non-zero elements of E; we shall write $F = (E, \theta, \Delta)$. In addition, we require that the involution be the natural one, i. e., $(\alpha + \beta u + \gamma v + \delta uv)^* = \alpha - \beta u - \gamma v - \delta uv$, with α, β, γ, δ in E. In fact, there exist other involutions in some quaternion division algebras relative to which no such basis representations exist: for example, the algebra $(Q_3, -3, -3)$, Q_p denoting the ordinary p-adic numbers ([1], p. 17), can easily be shown to be a division algebra by the argument of Section 6, below, together with the fact that the 3-adic Hilbert symbol $(-3, -3)$ is -1 ([7], pp. 26-27); now the map $\phi(\alpha + \beta u + \gamma v + \delta uv) = \alpha + \gamma u + \beta v + \delta uv$ is an anti-automorphic involution of $(Q_3, -3, -3)$, but the subspace of vectors a in $(Q_3, -3, -3)$ for which $\phi(a) = -a$ is only one-dimensional. The need for such a basis will be seen in § 3, below.

In the non-commutative case, $\operatorname{char} F \neq 2$ implies E is precisely the fixed subfield under the involution. In the commutative case, we *define* E to be this fixed subfield; thus F is a quadratic extension $E(\sqrt{\theta})$ of E, θ a non-square in E, and $(\alpha + \beta\sqrt{\theta})^* = \alpha - \beta\sqrt{\theta}$ with α, β in E.

Notations. Greek letters $\alpha, \beta, \cdots$ will denote elements known to belong to E; one of these is the *norm* $Na = aa^*$ of any a in F, specifically: $N(\alpha + \beta\sqrt{\theta}) = \alpha^2 - \beta^2\theta$ if F is commutative, $N(\alpha + \beta u + \gamma v + \delta uv) = \alpha^2 - \beta^2\theta - \gamma^2\Delta + \delta^2\theta\Delta$ if non-commutative. Let p denote a prime element (generator of the value group) for F. Let $\mathfrak{O} = \{a \in F \mid |a| \leqq 1\}$, the ring of *integers* of F; $\mathfrak{o} = \mathfrak{O} \cap E$, the integers of E. Let $\mathfrak{P} = \{a \in F \mid |a| < 1\}$, the unique 2-sided maximal prime ideal in $\mathfrak{O}$, $\mathfrak{p} = \mathfrak{P} \cap E$, the corresponding ideal in $\mathfrak{o}$. By assumption, the *residue class field* $\mathfrak{O}/\mathfrak{P}$ is finite, hence commutative ([2], Theorem 1.14). It should be noted that the proper $\mathfrak{O}$-ideals of F are all 2-sided and principal, and are precisely of the form $p^i\mathfrak{O} = \{a \in F \mid |a| \leqq |p|^i\}$, i running over all rational integers; they satisfy the strict inclusions $0 \subset \cdots \subset p^2\mathfrak{O} \subset \mathfrak{P} = p\mathfrak{O} \subset \mathfrak{O} \subset p^{-1}\mathfrak{O} \subset \cdots \subset F$. If $|a| = |p|^i$, i is called the *ordinal* of a.

The subfield E is also a (commutative) local field. To see this, note first that F is algebraic over E (for quaternion algebras, a is a zero of $x^2 - (a + a^*)x + aa^*$), so the restriction of $|\ |$ to E is non-trivial, by [11], Lemma 13, p. 18, and is of course non-archimedean and discrete. The residue class field $\mathfrak{o}/\mathfrak{p}$ of E is clearly finite. Completeness in the non-commutative case follows immediately from the fact that a quaternion algebra is by definition—or computation—central. If F is commutative, we note that the real-valued function $|\ |_1$ on F, given by $|a|_1 = |a^*|$ for a in F, defines a new valuation with respect to which F is complete, and so must be equivalent to $|\ |$, by [11], Lemma 18, p. 217; this means that the involution $a \rightarrow a^*$ is continuous, and so the fixed field E is here, too, complete under the restriction of $|\ |$.

Conversely, if we start with a commutative local field E and construct either a commutative quadratic extension or a quaternion division algebra F over E, Theorems 9 and 10 of [11], pp. 53-54, show that there is a unique extension of the valuation $|\ |$ on E up to F, given by $|a| = |Na|^{\frac{1}{2}}$; the residue class field $\mathfrak{O}/\mathfrak{P}$ is finite by Corollary 1 of [11], p. 20, and F, as a normed linear space over E, is complete under the extension ([1], pp. 18-20); thus F is a local field. In particular, this proves the *existence* of a non-commutative local field with involution, and shows $|a^*| = |a|$ in all cases.

We shall consider further questions relating to the arithmetic of the field extension, such as ramification, in § 5.

2. Hermitian spaces and lattices.

From the point of view of notation, the theory of equivalence of forms is considerably simplified if expressed in the language of geometric algebra. In this paragraph, we outline briefly these geometric concepts and their connection with hermitian forms—for further details, the reader is referred to [3] and to Chap. III of [2].

Definition. A *hermitian* (or *unitary*) *space* is a finite-dimensional vector space V over F with a scalar product from $V \times V$ into F, written $x \cdot y$, that is conjugate-symmetric ($x \cdot y = (y \cdot x)^*$) and F-linear in the first factor. By a *lattice in* V we shall mean a subset L of V that is also an $\mathfrak{O}$-module of finite rank relative to the module-structure on L inherited from F by $\mathfrak{O}$.

Of course, by "finite rank" we mean that L possesses a finite $\mathfrak{O}$-basis. Since all $\mathfrak{O}$-ideals are principal, any sub-$\mathfrak{O}$-module K of a lattice L is again a lattice, and in fact there exist bases $\langle x_1, \cdots, x_n \rangle$ for L and $\langle y_1, \cdots, y_m \rangle$ for K ($m \leqq n$) such that

$$y_\lambda = a_\lambda x_\lambda \qquad \text{for } 1 \leqq \lambda \leqq m, \tag{2.1}$$

the a_λ certain elements of $\mathfrak{O}$ (proven as in § 134 of [12]).

We will usually speak of a "lattice L" without explicit mention of the underlying space V, since it is the subspace $FL=\{ax \mid a\in F, x\in L\}$ of V that is actually of significance for us.

To indicate that hermitian spaces V and W are isometric ([2], Definition 3.2, p. 107), we shall write $V\simeq W$; similarly, a lattice-isometry (an $\mathfrak{O}$-linear product-preserving isomorphism) will be written $L\cong K$. Both types of isometry are equivalence relations. Clearly, $L\cong K$ implies $FL\simeq FK$, but not conversely: take $L=\mathfrak{O}x$ and $K=\mathfrak{O}(px)$, with $x^2=1$ (x^2 abbreviates $x\cdot x$).

To any n-ary form $\sum\sum\xi_\lambda a_{\lambda\mu}\xi_\mu^*$ there corresponds a hermitian space V, an abstract n-dimensional vector space with basis $\langle x_1,\cdots,x_n\rangle$ and scalar product defined by $x_\lambda\cdot x_\mu=a_{\lambda\mu}$, and also the lattice L in V defined by $L=\mathfrak{O}x_1+\cdots+\mathfrak{O}x_n$. Conversely, given an n-dimensional hermitian space V [or lattice L of rank n], there corresponds a hermitian form $\sum\sum\xi_\lambda(x_\lambda\cdot x_\mu)\xi_\mu^*$ defined by choosing a basis $\langle x_1,\cdots,x_n\rangle$ for V [respectively L]. Standard computations show that two forms are equivalent if and only if the corresponding spaces are isometric, *integrally* equivalent if and only if the corresponding lattices are isometric. The study of forms is thus replaced by that of spaces and lattices. Let us note here the following facts:

1) If V and W are isometric spaces, with bases $\langle x_\lambda\rangle$ and $\langle y_\lambda\rangle$ respectively, then there exists an invertible matrix T satisfying

$$((y_\lambda\cdot y_\mu))=T((x_\lambda\cdot x_\mu)){}^tT^*; \tag{2.2}$$

2) If L and K are isometric lattices, with bases $\langle x_\lambda\rangle$ and $\langle y_\lambda\rangle$ respectively, then there exists a unimodular matrix T satisfying equation (2.2).

Since $(x\cdot x)^*=x\cdot x$ for any vector x in V [or L], we see that x^2 always lies in the subfield E; such a scalar is said to be *represented* by V [L], or by x. A non-zero vector is called *isotropic* if it represents 0; following [10], we call a space or lattice *isotropic* if it contains an isotropic vector, otherwise *anisotropic*.

To indicate that a space V is the direct sum of mutually orthogonal subspaces $V_1,\cdots,V_t$, we shall write $V=V_1\oplus\cdots\oplus V_t=\bigoplus_{1\leqq\lambda\leqq t}V_\lambda$, and call each V_λ a *component* of V; the same notation will be used for lattices. On the other hand, given unrelated spaces $V_1,\cdots,V_t$, it is obvious how to construct a space $V=\bigoplus_{1\leqq\lambda\leqq t}W_\lambda$ with each $W_\lambda\simeq V_\lambda$, and we may write $V\simeq\bigoplus_{1\leqq\lambda\leqq t}V_\lambda$; similarly for lattices. Formula (2) on p. 265 of [6] shows that a *space* always has an *orthogonal* basis, in fact one containing a prescribed non-zero, non-isotropic vector; as shall be evident later, this need not be true for a lattice.

The orthogonal complement ([3], Definition 6, p. 12) in the space V [or lattice L] of a subset S will be denoted by $S^{\perp}$. The radical, $V^{\perp}$, will be written as $\operatorname{Rad} V$; as usual, V is termed *nondegenerate* if $\operatorname{Rad} V = 0$, otherwise *degenerate*. Note that $V = V_1 \oplus V_2$ is a nondegenerate space if and only if both V_1 and V_2 are nondegenerate as subspaces. For example, a 1-dimensional space or lattice cannot be both nondegenerate and isotropic; the *hyperbolic plane,* by definition a space with basis $\langle x, y\rangle$ satisfying $x^2 = y^2 = 0$, $x \cdot y = 1$, is an example of a 2-dimensional nondegenerate isotropic space (and, up to isometry, is the only such—cf. (3.1), below).

Every space V can be written $V = V_1 \oplus \operatorname{Rad} V$, V_1 any supplementary space to $\operatorname{Rad} V$; similarly, using (2.1), we may write $L = L_1 \oplus \operatorname{Rad} L$ for a lattice L; V_1 and L_1 are necessarily nondegenerate. As in [2], p. 116, there are canonical isometries $V_1 \simeq V/\operatorname{Rad} V$ and $L_1 \cong L/\operatorname{Rad} L$; since isometry of radicals is equivalent to equality of dimension [rank], it follows that we need only find isometry invariants for nondegenerate spaces and lattices. In the sequel, all spaces and lattices will be assumed *nondegenerate* and *non-zero.* As a consequence, $V = V_1 \oplus V_2$ implies $V_2 = V_1{}^{\perp}$, whereas $W^{\perp\perp} = W$ for any nondegenerate subspace W of a space V; the same is true for lattices.

Determinants. We assume here that F is *commutative.* Let V be a hermitian space, $\langle x_\lambda\rangle$ any basis, and consider $\det((x_\lambda \cdot x_\mu))$: if $\langle y_\lambda\rangle$ is any second basis, then by equation (2.2), $\det((x_\lambda \cdot x_\mu))$ and $\det((y_\lambda \cdot y_\mu))$ differ by a factor from $N(\dot{F})$ ($\dot{F}$ denoting the non-zero elements of F), and so the scalar $\det((x_\lambda \cdot x_\mu))$ is well-defined modulo $N(\dot{F})$; we use the notation $dV \simeq \det((x_\lambda \cdot x_\mu))$, and call dV the *determinant* of V. Similarly, for a lattice L with basis $\langle x_\lambda\rangle$, $\det((x_\lambda \cdot x_\mu))$ is well-defined modulo $N(\mathfrak{U})$, where $\mathfrak{U} = \{a \in F \mid |a| = 1\}$, the *units* of F, so we may define the *determinant* of L: $dL \cong \det((x_\lambda \cdot x_\mu))$. The following properties of dV and dL should be noted:

(2.3) $V \simeq W$ implies $dV \simeq dW$; similarly for lattices.

(2.4) If $V = V_1 \oplus V_2$, then $dV \simeq dV_1 \cdot dV_2$; similarly for lattices.

(2.5) If L, K are lattices, then $dL \cong dK$ implies $d(FL) \simeq d(FK)$, but not conversely.

(2.6) If L_1 is a sublattice of L such that $\operatorname{rank} L_1 = \operatorname{rank} L$ and $|dL_1| = |dL|$, then $L_1 = L$.

For example, (2.6) follows from (2.1); the other properties are obvious.

3. Fractional equivalence. We now quickly derive the invariants for space-isometry. Let V be a hermitian space. As shown in [6], pp. 266-267, the definition $x \# y = \frac{1}{2}[(x \cdot y) + (y \cdot x)]$ makes V into a quadratic space $V^{\#}$ over the subfield E (this is the reason for our requirement in § 1 that $a \rightarrow a^{*}$ be the "natural" involution in the non-commutative case), with $\dim V^{\#} = 2 \dim V$ if F is commutative, $\dim V^{\#} = 4 \dim V$ if F is a quaternion algebra. Note that $x \# x = x \cdot x$ for any x in V, and recall ([10], Remark 63:18) that a quadratic space of dimension at least 4 over a commutative local field field represents every non-zero scalar; this implies that a hermitian space V always represents 1 in the non-commutative case, and also in the commutative case provided $\dim V \geqq 2$. Thus $V = Fx_1 \oplus \cdots \oplus Fx_n$, with each $x_\lambda{}^2 = 1$ in the non-commutative case, and with $x_\lambda{}^2 = 1$ for $\lambda \leqq n - 1$ and $x_n{}^2 \cong dV$ in the commutative case; we therefore have the following result of Jacobson's:

THEOREM 3.1. *Let V and W be hermitian spaces over F. Then*

1) *if F is commutative, $V \cong W$ if and only if* $\dim V = \dim W$ *and* $dV \cong dW$;

2) *if F is a quaternion algebra, $V \cong W$ if and only if* $\dim V = \dim W$.

The following fact, proven similarly to its quadratic analogue ([10], Proposition 42:9), will be of considerable use to us:

(3.1) Let V be a 2-dimensional space. If V is isotropic, then it is a hyperbolic plane. If F is commutative, then V is isotropic if and only if $dV \cong -1$.

4. General properties of Hermitian lattices. Most of the ideas of this paragraph occur in the study of quadratic lattices—cf. [8] and [9]—and proofs will often be omitted. In the sequel, L and K will denote lattices.

If L has an orthogonal basis $\langle x_1, \cdots, x_n \rangle$, with $x_\lambda{}^2 = \alpha_\lambda$, we may often write

$$L = \mathfrak{O}x_1 \oplus \cdots \oplus \mathfrak{O}x_n \cong (\alpha_1) \oplus \cdots \oplus (\alpha_n),$$

or simply $L \cong (\alpha_1) \oplus \cdots \oplus (\alpha_n)$. If L is a lattice of rank 2 with basis $\langle x, y \rangle$ for which $x^2 = \alpha$, $y^2 = \beta$, and $x \cdot y = a$, we often write $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} \alpha & a \\ & \beta \end{pmatrix}$, or simply $L \cong \begin{pmatrix} \alpha & a \\ & \beta \end{pmatrix}$; a more complete notation would be $L \cong \begin{pmatrix} \alpha & a \\ b & \beta \end{pmatrix}$, where $b = y \cdot x$—but $b = a^{*}$, and so is superfluous. Lattices of rank 1 and 2 will be called *lines* and *planes*, respectively. In general, the notation $L = \mathfrak{O}x_1 + \cdots \mathfrak{O}x_n$ will be used only if $\langle x_1, \cdots x_n \rangle$ is a basis for L.

Definition. For any non-zero scalar a, define $aL = \{ax \mid x \in L\}$, also a lattice (in the space FL). Call a vector x of L *maximal in* L if x does not lie in pL.

PROPOSITION 4.1. a) *Each vector in a basis for L is maximal in L.*

b) *A maximal vector x in L can always be extended to a basis for L.*

c) *If x is any vector in L, there exists an integer a and a maximal vector y such that $x = ay$.*

This result is merely a statement of certain facts from elementary divisor theory of modules (cf. [10], § 81), and the proof is left to the reader. In particular, if the vector x in c) is isotropic, then so is y, which is therefore called a *maximal isotropic vector.*

Let α be any non-zero element of the subfield E. We define the *scaled lattice* $\alpha \circ L$ to be the *module* L with a new scalar product given by $x \circ y = \alpha(x \cdot y)$, where $x \cdot y$ denotes the old scalar product in L. (Note that the scalar-product laws need not be satisfied by $a \circ L$ if a does not lie in E.) The reader should be careful to distinguish between αL and $\alpha \circ L$.

One speaks of *scaling a vector,* which describes replacement of a vector x by bx, $|b| = 1$; for our purposes, this will always mean replacement of a basis $\langle x, y\rangle$ for a plane, with $|x \cdot y| = |p|^i$, by either $\langle p^i(x \cdot y)^{-1}x, y\rangle$ or $\langle x, (p^*)^i(y \cdot x)^{-1}y\rangle$, so as to obtain a new basis $\langle x_1, y_1\rangle$ with $x_1 \cdot y_1$ equal to p^i. In the sequel, we will understand "scaling" a vector to denote this operation, without further specification.

Definition. Let $sL = \{x \cdot y \mid x, y \in L\}$, $nL =$ the ideal generated by all scalars x^2 with x in L.

sL is an $\mathfrak{O}$-ideal, and measures the size of products of vectors of L; it is called the *scale* of L. nL measures the maximal size of squares of vectors; it is called the *norm* ideal of L in [8], but we shall not generally use this term, since it conflicts with the field norm. Clearly, $nL \subseteq sL$; if $nL = sL$, we call L *normal,* otherwise *subnormal.* For example, a line must be normal. The following properties of nL and sL should be noted:

(4.1) $s(aL) = a^2 sL$ and $n(aL) = a^2 nL$ for all non-zero a in F.

(4.2) $s(\alpha \circ L) = \alpha sL$ and $n(\alpha \circ L) = \alpha nL$ for all non-zero α in E.

(4.3) $s(L \oplus K) = sL + sK$ and $n(L \oplus K) = nL + nK$.

Definition. Let i be any ordinal. L will be called *p^i-modular* if $x \cdot L = p^i\mathfrak{O}$ for every maximal vector x in L.

Example of p^i-modular lattices are lines $\mathfrak{O}x$ with $|x^2| = |p|^i$, and planes $\mathfrak{O}x + \mathfrak{O}y$ with $|x \cdot y| = |p|^i$, $|x^2| < |p|^i$, $|y^2| < |p|^i$. As a special case of the latter, we define, for each ordinal i, the *hyperbolic plane* $H(i) \cong \begin{pmatrix} 0 & p^i \\ & 0 \end{pmatrix}$; by (3.1), the *spaces* $FH(i)$ are all isometric.

If L is p^i-modular, then $sL = p^i\mathfrak{O}$. Furthermore:

(4.4) If L is p^i-modular, then p^jL is p^{i+2j}-modular for any ordinal j, and $\alpha \circ L$ is p^{i+j}-modular if $|\alpha| = |p|^j$.

(4.5) $L_1 \oplus L_2$ is p^i-modular if and only if both L_1 and L_2 are.

By direct computation we can show:

PROPOSITION 4.2. *Suppose* $L = \mathfrak{O}x_1 + \cdots + \mathfrak{O}x_n$. *If the scalars* $a_\lambda = -\dfrac{x_\lambda \cdot x_1}{x_1^2}$ *are integral for all* $\lambda \geqq 2$, *then* $\mathfrak{O}x_1$ *is a component of* L, *and* $(\mathfrak{O}x_1)^\perp = \sum_{\lambda=2}^{n} \mathfrak{O}(x_\lambda + a_\lambda x_1)$. *If the scalars*

$$b_\lambda = \frac{(x_\lambda \cdot x_1)x_2^2 - (x_\lambda \cdot x_2)(x_2 \cdot x_1)}{(x_2 \cdot x_1)(x_1 \cdot x_2) - x_1^2 x_2^2} \text{ and } c_\lambda = \frac{(x_\lambda \cdot x_2)x_1^2 - (x_\lambda \cdot x_1)(x_1 \cdot x_2)}{(x_2 \cdot x_1)(x_1 \cdot x_2) - x_1^2 x_2^2}$$

are integral for all $\lambda \geqq 3$, *then* $\mathfrak{O}x_1 + \mathfrak{O}x_2$ *is a component of* L, *and* $(\mathfrak{O}x_1 + \mathfrak{O}x_2)^\perp = \sum_{\lambda=3}^{n} \mathfrak{O}(x_\lambda + b_\lambda x_1 + c_\lambda x_2)$.

Given $L = \mathfrak{O}x_1 + \cdots + \mathfrak{O}x_n$, either some $|x_i^2|$ or, if not, then $|x_i \cdot x_j|$ $(i \neq j)$, is maximal in the set $\{|x_\lambda \cdot x_\mu|\}_{1 \leqq \lambda, \mu \leqq n}$; in the former case $\mathfrak{O}x_i$ is therefore a component, in the latter $\mathfrak{O}x_i + \mathfrak{O}x_j$, and these were seen to be modular. By repeating this procedure, we obtain

PROPOSITION 4.3. *L can always be written as an orthogonal sum of modular lines and planes.*

Not all lattices—for example, a subnormal plane (cf. (4.3))—can have orthogonal bases. We prove:

PROPOSITION 4.4. *The p^i-modular lattice L has an orthogonal basis if any of the following conditions holds:*

a) *L has odd rank.*

b) *L is normal.*

c) $i = 0$ *and F contains an element a with* $|a| = |a + a^*| = 1$.

Proof. a) $L = \mathfrak{O}x_1 + \cdots + \mathfrak{O}x_n$, n odd. If $n \geqq 5$, and if we assume the result true for rank 3, then we may (Proposition 4.3) write

$L = L_1 \oplus \cdots \oplus L_r$ with, say, L_1 a line, and repeatedly apply the case $n = 3$ in the obvious way. Thus it suffices to take $n = 3$: $L = \mathfrak{O}x_1 \oplus (\mathfrak{O}x_2 + \mathfrak{O}x_3)$; by Proposition 4.2, we may assume $|x_2^2| < |x_2 \cdot x_3|$ and $|x_3^2| < |x_2 \cdot x_3|$, hence $|x_1^2| = |p|^i$ and (by scaling) $x_2 \cdot x_3 = p^i$. The scalars $a = -x_3^2 p^{-i}$ and $b = (p^{-i} x_2^2 x_3^2 - (p^*)^i)/x_1^2$ satisfy $|a| \leqq |p|$ and $|b| = 1$, so the vectors $y_1 = x_1 + b^{-1}(ax_2 + x_3)$, $y_2 = x_1 + x_2$, $y_3 = x_3$ give $L = \mathfrak{O}y_1 \oplus (\mathfrak{O}y_2 + \mathfrak{O}y_3)$ with $|y_2^2| = |p|^i$; we then may use Proposition 4.2 to find an orthogonal basis for $\mathfrak{O}y_2 + \mathfrak{O}y_3$.

b) By part a), we may assume $L = \mathfrak{O}x_1 + \cdots + \mathfrak{O}x_n$ with n even. L contains a vector x with $|x|^2 = |p|^i$; by Proposition 4.2, $L = \mathfrak{O}x_1 \oplus (\mathfrak{O}x_1)^{\perp}$, and a) may now be applied to $(\mathfrak{O}x_1)^{\perp}$.

c) By Proposition 4.3, it suffices to assume $L = \mathfrak{O}x + \mathfrak{O}y$ with $|x^2| < 1$, $|y^2| < 1$, and (by scaling) $x \cdot y = 1$; but then $L = \mathfrak{O}(ax + y) + \mathfrak{O}y$ with $|(ax + y)^2| = 1$, and Proposition 4.2 may now be applied. q. e. d.

Definition. For any ordinal j, define $L_{(j)} = \{x \in L \mid x \cdot L \subseteq p^j \mathfrak{O}\}$.

$L_{(j)}$ is a sublattice of L having the following properties:

(4.6) $s(L_{(j)}) \subseteq p^j \mathfrak{O}$ and $(L \oplus K)_{(j)} = L_{(j)} \oplus K_{(j)}$.

(4.7) If K is a p^i-modular sublattice of L, then K is a component of L if and only if $K \subseteq L_{(i)}$.

(4.8) If F is commutative, and L is p^i-modular of rank n, then $|dL| = |p|^{in}$.

These are proven as follows: (4.7) as in Proposition 1 of [9], (4.8) by choosing bases for the separate components in Proposition 4.3.

Definition. A decomposition $\bigoplus_{1 \leqq \lambda \leqq t} L_\lambda$ will be called a *Jordan splitting* if the L_λ are all modular and satisfy $sL_1 \supset sL_2 \supset \cdots \supset sL_t$. Jordan splittings $L = \bigoplus_{1 \leqq \lambda \leqq t} L_\lambda$ and $K = \bigoplus_{1 \leqq \lambda \leqq T} K_\lambda$ will be said to be of the *same type* if $t = T$ and, for each λ, $sL_\lambda = sK_\lambda$, $\operatorname{rank} L_\lambda = \operatorname{rank} K_\lambda$, and L_λ and K_λ are both normal or both subnormal. Given a Jordan splitting $\bigoplus L_\lambda$, define the ordinal $s(\lambda)$ by $sL_\lambda = p^{s(\lambda)}\mathfrak{O}$; thus $s(1) < s(2) < \cdots < s(t)$.

By (4.5) and Proposition 4.3, L always has a Jordan splitting. It is canonical in the sense that two Jordan splittings of isometric lattices—in particular, two splittings for the same lattice—are always of the same type (cf. [10], Theorem 91:9).

Thus $s(\lambda) = s_L(\lambda)$ is an invariant of L, not depending on a particular

Jordan splitting. Similarly, we may use any splitting $\bigoplus_{1 \leqq \lambda \leqq t} L_\lambda$ to define the invariants

$$u_L(\lambda) = u(\lambda) : n(L_{s(\lambda)}) = p^{u(\lambda)}\mathfrak{O} \qquad \text{for } 1 \leqq \lambda \leqq t,$$
$$\mathfrak{f}_L(\lambda) = \mathfrak{f}(\lambda) = p^{u(\lambda)+u(\lambda+1)-2s(\lambda)}\mathfrak{O} \qquad \text{for } 1 \leqq \lambda \leqq t-1.$$

These have the following properties ($\bigoplus L_\lambda$ denotes an arbitrary Jordan splitting for L):

(4.9) $u(1) \leqq u(2) \leqq \cdots \leqq u(n)$.

(4.10) $nL_\lambda \subseteq p^{u(\lambda)}\mathfrak{O} \subseteq p^{s(\lambda)}\mathfrak{O}$ for each λ.

(4.11) If L has normal $p^{s(\lambda)}$-modular component, then $nL_\lambda = p^{u(\lambda)}\mathfrak{O}$.

(4.12) $\mathfrak{f}(\lambda) \subseteq p\mathfrak{O}$ for all λ.

(4.13) $s(\lambda)$, $u(\lambda)$, $\mathfrak{f}(\lambda)$ are invariant under isometry; $\mathfrak{f}(\lambda)$ is invariant under scaling (i.e., $\mathfrak{f}_L(\lambda) = \mathfrak{f}_{\alpha \circ L}(\lambda)$).

In view of (4.10), O'Meara calls a Jordan splitting $\bigoplus L_\lambda$ *saturated* if $nL_\lambda = p^{u(\lambda)}\mathfrak{O}$ for all λ. That every lattice has a saturated splitting can be shown as in Proposition 2 of [9], using the following concept:

Definition. $\mathfrak{R}L = \{x \in FL \mid x \cdot L \subseteq \mathfrak{O}\}$.

As in [9], p. 162, we can easily show that $\mathfrak{R}L$ is a lattice with the properties

(4.14) $\mathfrak{R}(L \oplus K) = \mathfrak{R}L \oplus \mathfrak{R}K$ and $\mathfrak{R}\mathfrak{R}L = L$.

(4.15) If L is p^i-modular, then $\mathfrak{R}L = p^{-i}L$ (thus p^{-i}-modular).

5. Ramification in the extension F/E.

From now on, π will denote a prime element for E (thus $\mathfrak{p} = \pi\mathfrak{o}$), $\mathfrak{u} = \{\alpha \in E \mid |\alpha| = 1\}$ the multiplicative group of *units* of E, $|F| = \{|a| \mid a \neq 0 \text{ and } a \in F\}$ the *value group* of F, $|E|$ the value group of E. We also borrow a notation from [8]: to indicate that an otherwise-unspecified scalar a has value $|a| \leqq |b|$, we shall write $a = \{b\}$; to indicate $|a| = |b|$, write $a = [b]$. Thus, $g = h + \{p^2\}$ means $|g - h| \leqq |p^2|$, $g = h + [\pi]$ means $|g - h| = |\pi|$ and $g - h$ lies in E.

Now, $|E| \subseteq |F|$ always; if $|E| = |F|$, the extension is called *unramified,* otherwise *ramified.* When E is commutative, Theorem 6 on p. 55 of [1] shows that the product of the group index $(|F| : |E|)$ and the residue class degree $[\mathfrak{O}/\mathfrak{P} : \mathfrak{o}/\mathfrak{p}]$ ($\mathfrak{o}/\mathfrak{p}$ is naturally embedded in $\mathfrak{O}/\mathfrak{P}$) is equal to 2; thus either

1) F/E unramified $\Rightarrow |F| = |E|$ and $\mathfrak{O}/\mathfrak{P}$ is a quadratic extension of $\mathfrak{o}/\mathfrak{p}$; or

2) F/E ramified $\Rightarrow |E|$ has index 2 in $|F|$ and $\mathfrak{O}/\mathfrak{P} = \mathfrak{o}/\mathfrak{p}$.

If $|2| = 1$, F is called *non-dyadic*; if $|2| < 1$, *dyadic*.

The next two facts describe a "critical point" for squares in a local field.

(5.1) The subfield E contains a non-square unit of the form $1 + [4]$ (cf [10], Proposition 63:4).

(5.2) $1 + \{4\pi\}$ is always a square in E (cf. [10], Theorem 63:1).

In the sequel, (5.2) will be referred to as "Hensel's lemma," although it is in fact only a special case of that lemma ([11], Theorem 4, p. 35).

Suppose F is commutative and dyadic. Since $\mathfrak{o}/\mathfrak{p}$ is finite of characteristic 2, it contains square roots of all elements; thus

(5.3) Given α in $\mathfrak{o}$, there exists β in $\mathfrak{o}$ such that $\beta^2 \equiv \alpha \pmod{\pi}$.

(5.3) will be used repeatedly below. In particular, it shows that (still in the commutative dyadic case) every E-unit α can be written as $\beta^2 + \gamma$ with $|\gamma| < 1$; the *smallest* E-ideal $\gamma\mathfrak{o}$ for which such a representation exists will be denoted by $\mathfrak{d}_E(\alpha)$. By Hensel's lemma, $\mathfrak{d}_E(\alpha)$ is either 0 (if α is a square in E) or else includes $4\mathfrak{o}$ (if a non-square); in fact, according to Propositions 63:2 and 63:5 of [10], the E-ideals occurring as $\mathfrak{d}_E(\alpha)$ are precisely 0, $4\mathfrak{o}$, and all $\pi^{2k+1}\mathfrak{o}$ for which $|4| < |\pi|^{2k+1} < 1$, and furthermore $\mathfrak{d}_E(1 + [\pi^{2k+1}])$ is precisely $\pi^{2k+1}\mathfrak{o}$ for such k.

In the commutative case, $F = E(\sqrt{\theta})$; possibly replacing θ by $\pi^m\theta$ for suitable ordinal m, we shall assume θ is either a (non-square) unit or a prime in E. Using Hensel's lemma and our observations about $\mathfrak{d}_E$, ramification is readily characterized as follows:

Non-dyadic Case. F/E is ramified if $|\theta| = |\pi|$, unramified if $|\theta| = 1$.

Dyadic Case. a) F/E is ramified if $|\theta| = |\pi|$.

b) F/E is ramified if $|\theta| = 1$ and $\mathfrak{d}_E(\theta) = \pi^{2k+1}\mathfrak{o}$ with $|4| < |\pi|^{2k+1} < 1$, and so $F = E((1 + \pi^{2k+1}\delta)^{\frac{1}{2}})$ with $|\delta| = 1$.

c) F/E is unramified if $|\theta| = 1$ and $\mathfrak{d}_E(\theta) = 4\mathfrak{o}$, and so $F = E((1 + 4\delta)^{\frac{1}{2}})$ with $|\delta| = 1$.

We will refer to ramified dyadic extensions of type a) as *ramified-prime* ("R-P"), those of type b) as *ramified-unit* ("R-U"); that a ramified extention cannot be both R-P and R-U is easily seen by expressing $(1+\pi^{2k+1}\delta)$ as $\alpha+\beta\sqrt{\pi}$ and squaring.

Examples. Non-dyadic case. $Q_3(\sqrt{3})/Q_3$ is ramified, $Q_3(\sqrt{2})/Q_3$ unramified.

Dyadic case. $Q_2(\sqrt{2})/Q_2$ is R-P, $Q_2(\sqrt{3})/Q_2$ is R-U, $Q_2(\sqrt{5})/Q_2$ is unramified.

6. The non-commutative case. In this paragraph, F denotes a quaternion division algebra over E, π a fixed prime element for E. Let us note immediately that ([10], Proposition 63:11b) E is necessarily isomorphic to the algebra (E,π,ϵ), ϵ denoting the special unit of (5.1). Moreover, since an isomorphism ϕ of quaternion algebras commutes with the basic involution, i.e., $(\phi x)^*=\phi(x^*)$ (cf. [10], Proposition 57:6), the scalar-product structure is preserved under such isomorphism, so we may assume $F=E+Eu+Ev+Euv$ with $u^2=\pi$, $v^2=\epsilon$. Since $|a|=|Na|^{\frac{1}{2}}$ for any a in F, F/E is ramified, with index $(|F|:|E|)=2$, and so $u=\sqrt{\pi}$ will serve as prime element for F; we therefore write $u=p=\sqrt{\pi}$, $v=\sqrt{\epsilon}$.

PROPOSITION 6.1. *Let L be a p^i-modular lattice of rank n. Then*

a) $L\cong(\pi^{i/2})\oplus(\pi^{i/2})\oplus\cdots\oplus(\pi^{i/2})$ *if i is even;*

b) $L\cong H(i)\oplus H(i)\oplus\cdots\oplus H(i)$ *if i is odd.*

Proof. Scaling L by a suitable power of π, we may assume $i=0$ or 1.

a) $i=0$. By Proposition 4.4, c), with $a=1$ in the non-dyadic case, $a=\frac{1}{2}(1+\sqrt{\epsilon})$ in the dyadic case, L has an orthogonal basis; it thus suffices to show that a line $\mathfrak{O}x$, with $|x^2|=1$, represents 1, and this is clear since Fx does (Theorem 3.1).

b) $i=1$. Since F/E is ramified, $sL\subseteq p^2\mathfrak{O}$, and so L is, by (4.5) and Proposition 4.3, an orthogonal sum of p-modular planes; it thus suffices to assume L is a plane. FL is isotropic (cf. § 3), so with x maximal isotropic in L, we may, after scaling, write $L=\mathfrak{O}x+\mathfrak{O}y\cong\begin{pmatrix}0 & p\\ & \{\pi\}\end{pmatrix}$; if we set $b=-\frac{1}{2}y^2p^{-1}$ in the non-dyadic case, $b=-\frac{1}{2}(1+\sqrt{\epsilon})y^2p^{-1}$ in the dyadic case, then b is an integer, so $L=\mathfrak{O}x+\mathfrak{O}(bx+y)\cong\begin{pmatrix}0 & p\\ & 0\end{pmatrix}\cong H(1)$.

q. e. d.

It follows that two p^i-modular lattices are isometric if and only if they have the same rank, and so we conclude

THEOREM 6.2. *Suppose F is a quaternion division algebra. Then two lattices are isometric if and only if they are of the same type.*

For the remainder of the paper, F will be assumed *commutative.*

7. The unramified case. Whether or not F/E is ramified, local class field theory ([11], Theorem 4, p. 163) shows that the norm group of F has index 2 in the group of non-zero elements of E. In the unramified case, we may take π to be a prime element for both F and E; since π cannot be a norm, this shows that every non-zero element of E having even ordinal is a norm, every element of odd ordinal a norm times π. Now, F satisfies c) of Proposition 4.4, with $a = 1$ in the non-dyadic case and $a = \frac{1}{2}(1 + (1 + 4\delta)^{\frac{1}{2}})$ in the dyadic case; thus *every* lattice has an orthogonal basis (scaling by any power p^i is allowed). Combining these observations, we see that any p^i-modular lattice may be written $L \cong (\pi^i) \oplus \cdots \oplus (\pi^i)$. As in the non-commutative case, we have

THEOREM 7.1. *Suppose F/E is unramified. Then two lattices are isometric if and only if they are of the same type.*

8. The ramified non-dyadic case. As in § 5, we shall take $F = E(\sqrt{\pi})$ and $p = \sqrt{\pi}$, π a suitably chosen prime of E. The following result should be compared with Propositions 6.1 and 10.4.

PROPOSITION 8.1. *Let L be a p^i-modular lattice of rank n. Then*

a) $L \cong (\pi^{i/2}) \oplus (\pi^{i/2}) \oplus \cdots \oplus (\pi^{i/2}) \oplus (\pi^{-(n-1)i/2} dL)$ if i is even;

b) $L \cong H(i) \oplus H(i) \oplus \cdots \oplus H(i)$ if i is odd.

Proof. As in Proposition 6.1, we may assume $i = 0$ or 1.

a) $i = 0$. By Proposition 4.4, c), with $a = 1$, any component of L has an orthogonal basis. Now, as is well-known (cf. [2], p. 143), a binary quadratic form over a finite field of characteristic not 2 represents any non-zero element of that field; thus each binary component of L represents 1 modulo π, with coefficients from $\mathfrak{o}$, and so by Hensel's lemma represents 1. Using Proposition 4.2, the result then follows by induction on n.

b) $i = 1$. As in Proposition 6.1, we may assume L is a subnormal plane, say (after suitable scaling) $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} \{\pi\} & p \\ & \{\pi\} \end{pmatrix}$; but then $d(FL) \simeq \{\pi^2\} - pp^* = -pp^*(1 + \{\pi\}) \simeq -1$ by Hensel's lemma, so ((3.1)) FL is isotropic, and the result follows as in Proposition 6.1. q. e. d.

It follows that two p^i-modular lattices are isometric if and only if they have the same rank and—when i is even—the same determinant.

THEOREM 8.2. *Suppose F/E is ramified non-dyadic. Then the lattices L and K are isometric if and only if both of the following conditions hold:*

1) *L and K are of the same type.*

2) *Given Jordan splittings $L = \bigoplus L_\lambda$ and $K = \bigoplus K_\lambda$, $dL_j \cong dK_j$ for each index j for which $s(j)$ is even.*

Proof. By what has been shown, it suffices to prove necessity of 2); we may assume $L = K$ and $s(j) = 0$. With ϕ denoting the projection map: $L_j \to K_j$, we easily see that $x \cdot y = (\phi x) \cdot (\phi y) + \{p\}$ for all x, y in L_j, in particular for vectors x, y of an orthogonal basis for L_j; thus by Hensel's lemma, $dL_j \cong d(\phi L_j)$, so $|d(\phi L_j)| = |dK_j|$ by (4.8). By nondegeneracy, ϕ is monomorphic, hence rank-preserving: $\operatorname{rank} \phi L_j = \operatorname{rank} L_j = \operatorname{rank} K_j$. By (2.6), $\phi L_j = K_j$, so $dL_j \cong dK_j$. q. e. d.

9. The ramified dyadic case. Hyperbolic planes. For the remainder of the paper, we assume F/E is ramified and $|2| < 1$. In view of § 5, we have, for some prime π in E, $F = E(\sqrt{\pi})$ in R-P, and shall take $p = \sqrt{\pi}$; $F = E((1 + {}^{2k+1}\delta)^{\frac{1}{2}})$ in R-U, with $|\delta| = 1$ and $|4| < |\pi|^{2k+1} < 1$, and shall take $p = \dfrac{1 + (1 + \pi^{2k+1}\delta)^{\frac{1}{2}}}{\pi^k}$ (note in fact that $|p| = \sqrt{|\pi|}$).

The non-negative ordinal k in R-U will be used repeatedly in computation, but will not appear among the final invariants. It is worth noting that $pp^* = -\pi\delta$ in R-U.

Expressing the arbitrary F-scalar $a = \alpha + \beta\sqrt{\theta}$, it will be convenient to call α the "real part" of a: $\alpha = \operatorname{Re} a$. The following estimate will be useful:

$$(9.1)\quad \begin{cases} \text{In R-P, with } i \text{ even}, & |2\operatorname{Re}\{p^i\}| \leqq |2p^i| \quad \text{and } nH(i) = 2p^i\mathfrak{D} \\ \text{R-P}, \quad i \text{ odd}, & |2\operatorname{Re}\{p^i\}| \leqq |2p^{i+1}| \quad \text{and } nH(i) = 2p^{i+1}\mathfrak{D} \\ \text{R-U}, \quad i \text{ even}, & |2\operatorname{Re}\{p^i\}| \leqq |2p^{i-2k}| \quad \text{and } nH(i) = 2p^{i-2k}\mathfrak{D} \\ \text{R-U}, \quad i \text{ odd}, & |2\operatorname{Re}\{p^i\}| \leqq |2p^{i-1-2k}| \quad \text{and } nH(i) = 2p^{i-1-2k}\mathfrak{D} \end{cases}$$

To prove the estimate on $|2\operatorname{Re}\{p^i\}|$, note that since any integer of F can be written $\alpha + \beta p$ with α, β in $\mathfrak{o}$ (1 and p form an integral basis for F/E), we may express $\{p^i\} = \pi^{i/2}(\alpha + \beta p)$ if i is even, $\{p^i\} = \pi^{(i-1)/2}p^*(\alpha + \beta p)$ if i is odd, and then $\operatorname{Re}\{p^i\}$ is easily computed. To compute $nH(i)$, we may assume, by scaling, that $i = 0$ or 1. Expressing $H(i) = \mathfrak{D}x + \mathfrak{D}y \cong \begin{pmatrix} 0 & p^i \\ & 0 \end{pmatrix}$, we have $(ax+by)^2 = 2\,\mathrm{Re}(ap^ib^*)$ for any integers a, b, so by the estimate for $|2\,\mathrm{Re}\{p^i\}|$, $nH(i)$ is surely included in the correct ideal; the reverse inclusion follows from the fact that this ideal is equal to $z^2\mathfrak{D}$, where z is the vector $x+y$ in R-P with $i=0$ or R-U with $i=1$, $px+y$ in R-P with $i=1$ or R-U with $i=0$. Moreover, it follows that $H(i)$ is always subnormal (recall $|2| \leqq |\pi|^{k+1}$ in R-U).

Proposition 9.1. a) *If L is p^i-modular, then $nL \supseteq nH(i)$.*

b) *If α is any E-scalar in $nH(i)$, then*

$$H(i) \cong \begin{pmatrix} 0 & p^i \\ & \alpha \end{pmatrix}.$$

Proof. By scaling, assume $i=0$ or 1.

a) It suffices to assume L a subnormal plane, say (cf. (9.1)) $L = \mathfrak{D}x + \mathfrak{D}y \cong \begin{pmatrix} \{\pi\} & p^i \\ & \{\pi\} \end{pmatrix}$. If either $x^2\mathfrak{D} \supseteq nH(i)$ or $y^2\mathfrak{D} \supseteq nH(i)$, we are done, and otherwise $nL \supseteq z^2\mathfrak{D} = nH(i)$ for the vector z used above to compute $nH(i)$.

b) Write $L = \mathfrak{D}x + \mathfrak{D}y \cong \begin{pmatrix} 0 & p^i \\ & 0 \end{pmatrix}$; by (9.1), we may write (β an element of $\mathfrak{o}$)

$$\alpha = \left\{ \begin{array}{ll} 2\beta & \text{in R-P, } i=0 \\ 2\pi\beta & \text{in R-P, } i=1 \\ 2\pi^{-k}\beta & \text{in R-U} \end{array} \right\},$$

$$\text{and } y_1 = \left\{ \begin{array}{ll} \beta x + y & \text{in R-P with } i=0\text{, R-U with } i=1 \\ p\beta x + y & \text{in R-P with } i=1\text{, R-U with } i=0 \end{array} \right\}$$

then gives $L = \mathfrak{D}x + \mathfrak{D}y_1 \cong \begin{pmatrix} 0 & p^i \\ & \alpha \end{pmatrix}$. q. e. d.

Proposition. 9.2. *Let L be a p^i-modular plane with $nL = nH(i)$.*

a) *If L is isotropic, then $L \cong H(i)$.*

b) *In* R-P *with i odd and in* R-U *with i even, L must be isotropic. In the other two cases, if L is anisotropic, then $x^2\mathfrak{D} = nL$ for any maximal vector x.*

c) *If K is another p^i-modular plane, with $nK = nL$ and $dK \cong dL$, then $K \cong L$.*

Proof. By scaling, assume $i = 0$ or 1.

a) $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} 0 & p^i \\ & \{\pi\} \end{pmatrix}$. Defining the scalar

$$a = \left\{\begin{matrix} -y^2/2 & \text{in R-P}, i = 0 \\ -y^2/2p & \text{in R-P}, i = 1 \\ -py^2/2\pi^{-k} & \text{in R-U}, i = 0 \\ -y^2/2\pi^{-k} & \text{in R-U}, i = 1 \end{matrix}\right\},$$

we have $L = \mathfrak{O}x + \mathfrak{O}(ax + y) \cong \begin{pmatrix} 0 & p^i \\ & 0 \end{pmatrix} \cong H(i)$.

b) *R-P case.* If $i = 1$, $L \cong \begin{pmatrix} \{2\pi\} & p \\ & \{2\pi\} \end{pmatrix}$, so

$$d(FL) \simeq -pp^*(1 + \{4\pi\}) \simeq -1$$

(Hensel's lemma); thus L is isotropic, by (3.1). If $i = 0$, then $|x^2| < |2|$ implies $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} \{2\pi\} & 1 \\ & \{2\} \end{pmatrix}$, $d(FL) \simeq -1$, L again isotropic, so that anisotropy requires $x^2\mathfrak{O} = 2\mathfrak{O} = nH(0)$.

R-U case. Write $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} [2\pi^{-k+r}] & p^i \\ & 2\pi^{-k+s}\alpha \end{pmatrix}$ with $r \geqq 0$, $s \geqq 0$, $|\alpha| = 1$. First suppose $i = 0$; setting $y_1 = -\alpha p\pi^s x + y$, we have $y_1^2 = \{2\pi^{-k+s+1}\}$ and therefore $L = \mathfrak{O}x + \mathfrak{O}y_1 \cong \begin{pmatrix} [2\pi^{-k+r}] & p^i \\ & \{2\pi^{-k+s+1}\} \end{pmatrix}$, and repetition of this procedure eventually shows, as in R-P, that L is isotropic. Next suppose $i = 1$; we must show $r = 0$, so assume $r \geqq 1$. Setting $y_1 = -\alpha\pi^s x + y$, we have $y_1^2 = \{2\pi^{-k+s+1}\}$ and again $L \cong \begin{pmatrix} [2\pi^{-k+r}] & p^i \\ & \{2\pi^{-k+s+1}\} \end{pmatrix}$, isotropic; anisotropy therefore requires $r = 0$, i.e., $x^2 = |2\pi^{-k}|$, $x^2\mathfrak{O} = nH(1)$.

c) Note that $dL \cong dK$ requires, by (3.1), that L and K must be both isotropic or both anistropic; by a), we need only consider the latter case. By b), write $L = \mathfrak{O}x_1 + \mathfrak{O}y_1$ and $K = \mathfrak{O}x_2 + \mathfrak{O}y_2$ with

$$x_1^2\mathfrak{O} = y_1^2\mathfrak{O} = x_2^2\mathfrak{O} = y_2^2\mathfrak{O} = nH(i) \text{ and } x_1 \cdot y_1 = x_2 \cdot y_2 = p^i.$$

Since the *spaces* FL and FK are isometric, we may assume $FL = FK$, and can thus write $x_1 = a_{11}x_2 + a_{12}y_2$, $y_1 = a_{21}x_2 + a_{22}y_2$, the $a_{\lambda\mu}$ in F; in fact, by a) of Proposition 4.1 and b), above, the $a_{\lambda\mu}$ must all be integers. Thus $L \subseteq K$, and the result follows by (2.6). q. e. d.

As with quadratic forms, hyperbolic planes may be cancelled from isometric lattices. Specifically:

PROPOSITION 9.3. *Suppose* $L_1 \oplus H_1 \cong L_2 \oplus H_2$, *with* $H_1 \cong H_2 \cong H(i)$ *for some ordinal* i. *Then* $L_1 \cong L_2$.

We do not give a proof, since it is similar to the analogous result for quadratic forms ([9], p. 165), with the usual exception that we must consider p^0- and p-modular lattices separately. Note also that Proposition 1 of [9] is our (4.7), while Theorem 14.3 of [8] may be replaced by our Proposition 9.2, c).

The general cancellation law—$L \oplus K \cong J \oplus K$ implies $L \cong J$ for arbitrary K—is not necessarily true in the ramified dyadic case: for a counterexample (cf. [5], p. 696), take $E = Q_2$, $F = Q_2(\sqrt{2})$, $L \cong (1) \oplus (-1)$, $K \cong (1)$, $J \cong \begin{pmatrix} 0 & -1 \\ & 0 \end{pmatrix} \cong H(0)$. In the earlier cases—non-commutative, unramified, and ramified-non-dyadic—full cancellation is valid, as can easily be seen by considering the isometry invariants we have computed.

10. The ramified dyadic case. Modular lattices. Having studied p^i-modular planes L with nL minimal $(= nH(i))$, we next consider those with $nL \supset nH(i)$. For quadratic forms ([8], Theorem 14.3), two such planes are isometric provided they have the same determinant and represent the same numbers. The next result—the proof of which depends on the presence of an extension field F from which to choose suitable scalars—shows that, in the hermitian case, "represent the same numbers" may be replaced by the weaker "have the same norm ideal"; as a consequence, the final set of invariants is simpler in the hermitian than the quadratic case (cf. Theorem 11.4, below, and Theorem 2 of [9]).

PROPOSITION 10.1. *Let* L *be a* p^i*-modular plane with* $nL = p^{2m}\mathfrak{O} \supset nH(i)$. *Then* L *represents every non-zero element of* nL.

Proof. Suppose we have proven that L represents π^m, and are given arbitrary α in $p^{2m}\mathfrak{O}$, say $\alpha = \pi^{m+r}\beta$ with $r \geqq 0$ and $|\beta| = 1$: then if r is even, the lattice $\beta^{-1} \circ L$ will represent π^m, so L contains a vector x with $x^2 = \pi^m\beta$, and $(\pi^{r/2}x)^2 = \alpha$; if r is odd, then $(\pi\beta/pp^*)^{-1} \circ L$ represents π^m, L contains x with $x^2 = \pi^{m+1}\beta/pp^*$, and $(\pi^{(r-1)/2}px)^2 = \alpha$.

It thus suffices to show L represents π^m; by scaling, assume $i = 0$ or 1. Now, L surely represents $\pi^m + \{\pi^{m+1}\}$, since it contains a vector z with $z^2 = [\pi^m]$, and if we pick α in $\mathfrak{u}$ with $\alpha^2 \equiv \pi^m/z^2 \pmod{\pi}$, then $(\alpha z)^2 = \pi^m + \{\pi^{m+1}\}$. Our procedure will be to assume L contains a vector x with $x^2 = \pi^m + [\pi^{m+r}]$, $r \geqq 1$, and then to show L also represents $\pi^m + \{\pi^{m+r+1}\}$;

by induction on r, Hensel's lemma can be invoked to find a representation of π^m. There are several cases.

1) *L normal.* Here $i = m = 0$, $|x^2| = 1$, so $L = \mathfrak{O}x \oplus \mathfrak{O}y$ with $|y^2| = 1$. For r even, pick β in $\mathfrak{u}$ with $\beta^2 \equiv (1 - x^2)/\pi^r y^2 \pmod{\pi}$ and set $x_1 = x + \beta\pi^{r/2}y$; for r odd, pick β in $\mathfrak{u}$ with $\beta^2 \equiv (1 - x^2)/pp^*\pi^{r-1}y^2 \pmod{\pi}$ and set $x_1 = x + \beta p\pi^{(r-1)/2}y$. In either case, $x_1{}^2 = 1 + \{\pi^{r+1}\}$.

Suppose then L is subnormal; write $L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} \pi^m + [\pi^{m+r}] & p^i \\ & \{\pi^m\} \end{pmatrix}$

2) *R-P, $i = 1$ with $nL = 2\mathfrak{O}$.* We first express $L = \mathfrak{O}x + \mathfrak{O}y_1 \cong \begin{pmatrix} x^2 & p \\ & [\pi^m] \end{pmatrix}$, permissible because if $|y^2| < |\pi|^m$, then $y_1 = x + y$ will work (after suitable scaling). Pick β in $\mathfrak{u}$ with $\beta^2 \equiv -y_1{}^2/x^2 \pmod{\pi}$, set $z = \beta x + y_1$ and $x_1 = x + \frac{1}{2}(\pi^m - x^2)(z \cdot x)^{-1}z$ (note that $\frac{1}{2}(\pi^m - x^2)(z \cdot x)^{-1}$ is an integer); then since $|z^2| \leqq |\pi|^{m+1}$, we have $x_1{}^2 = \pi^m + \{\pi^{m+r+1}\}$.

3) *R-P, $i = 1$ with $nL \supset 2\mathfrak{O}$, or $i = 0$.* Here we seek expressions $L = \mathfrak{O}x + \mathfrak{O}y_1 \cong \begin{pmatrix} x^2 & p^i \\ & [\pi^m] \end{pmatrix}$ and $L = \mathfrak{O}x + \mathfrak{O}y_2 \cong \begin{pmatrix} x^2 & p^i \\ & [\pi^{m+1}] \end{pmatrix}$: the former is handled as in case 2); for the latter, pick γ in $\mathfrak{u}$ with $\gamma^2 \equiv -x^2/y_1{}^2 \pmod{\pi}$ and set $y_2' = x + \gamma y_1$, which will (after scalling) work as our y_2 unless $|(y_2')^2| \leqq |\pi|^{m+2}$, in which case $px + y_2'$ will work. Now choose the unit β and vector x as in the following table, the choice depending on the value of i and parity of r:

(i, r)	$(0, \text{even})$	$(0, \text{odd})$	$(1, \text{even})$	$(1, \text{odd})$
$\beta^2 \pmod{\pi}$	$\equiv \dfrac{\pi^m - x^2}{pp^*\pi^{r-2}y_2{}^2}$	$\equiv \dfrac{\pi^m - x^2}{pp^*\pi^{r-1}y_1{}^2}$	$\equiv \dfrac{\pi^m - x^2}{\pi^r y_1{}^2}$	$\equiv \dfrac{\pi^m - x^2}{\pi^{r-1}y_2{}^2}$
x_1	$x + \beta p\pi^{(r-2)/2}y_2$	$x + \beta p\pi^{(r-1)/2}y_1$	$x + \beta\pi^{r/2}y_1$	$x + \beta\pi^{(r-1)/2}y_2$

In all cases, an easy computation shows $x_1{}^2 = \pi^m + \{\pi^{m+r+1}\}$.

4) *R-U,* $|\pi|^{m+r} > |2\pi^{-k+i}|$. The preliminary expression here is $L = \mathfrak{O}x + \mathfrak{O}y_1 \cong \begin{pmatrix} x^2 & p^i \\ & [\pi^m] \end{pmatrix}$, obtained by taking $y_1 = x + y$ if $|y^2| < |\pi|^m$. Now if r is even, pick β in $\mathfrak{u}$ with $\beta^2 \equiv (\pi^m - x^2)/\pi^r y_1{}^2 \pmod{\pi}$ and set $x_1 = x + \beta\pi^{r/2}y_1$; if r is odd, pick β in $\mathfrak{u}$ with $\beta^2 \equiv (\pi^m - x^2)/pp^*\pi^{r-1}y_1{}^2 \pmod{\pi}$ and set $x_1 = x + \beta p\pi^{(r-1)/2}y_1$. Again, $x_1{}^2 = \pi^m + \{\pi^{m+r+1}\}$.

5) *R-U,* $|\pi|^{m+r} \leqq |2\pi^{-k+i}|$. This time we seek to express

$$L = \mathfrak{O}x + \mathfrak{O}y_1 \cong \begin{pmatrix} x^2 & p^i \\ & \{2\pi^{-k+i}\} \end{pmatrix} \text{ given, say, } L = \mathfrak{O}x + \mathfrak{O}y \cong \begin{pmatrix} x^2 & p^i \\ & [\pi^{m+s}] \end{pmatrix}$$

with non-negative ordinal s, and we proceed as follows: if s is even, pick β in $\mathfrak{u}$ with $\beta^2 \equiv -y^2/\pi^s x^2 \pmod{\pi}$ and set $y' = \beta\pi^{s/2}x + y$, while if s is odd, pick β in $\mathfrak{u}$ with $\beta^2 \equiv -y^2/pp^*\pi^{s-1}x^2 \pmod{\pi}$ and set $y = \beta p^*\pi^{(s-1)/2}x + y$; provided $|\pi|^{m+s} > |2\pi^{-k+i}|$, this procedure, which yields $L = \mathfrak{O}x + \mathfrak{O}y'$ with $|(y')^2| \leqq |\pi|^{m+s+1}$ (for the case $i = 1$ and $s = 0$, cf. (9.1)), may be repeated until the desired expression is obtained. Once we have found y_1, the vector $x_1 = x + \frac{1}{2}\pi^k(\pi^m - x^2)(p^*)^{1-i}y_1$ is seen to satisfy $x_1{}^2 = \pi^m + \{\pi^{m+r+1}\}$.

q. e. d.

PROPOSITION 10.2. *Let L be a $\mathfrak{p}^i$-modular plane, $i = 0$ or 1, with $\mathfrak{n}L = \mathfrak{p}^{2m}\mathfrak{O} \supset \mathfrak{n}H(i)$.*

a) *If L is normal, then $L \cong (1) \oplus (dL) \cong \begin{pmatrix} 1 & 1 \\ & \{1\} \end{pmatrix}$.*

b) *If L is subnormal, then $L \cong \begin{pmatrix} \pi^m & p^i \\ & \{\alpha\} \end{pmatrix}$, where $\alpha = 4\pi^{-m+i}$ in R-P, $\alpha = 4\pi^{-2k-m+i-1}$ in R-U.*

Proof. By Proposition 10.1, L contains a vector x with $x^2 = \pi^m$. If L is normal, then $L = \mathfrak{O}x \oplus \mathfrak{O}y$ with $y^2 \cong dL$, and also $L = \mathfrak{O}x + \mathfrak{O}(x+y)$ $\cong \begin{pmatrix} 1 & 1 \\ & \{1\} \end{pmatrix}$. Next, suppose L is subnormal, and write $L = \mathfrak{O}x + \mathfrak{O}y$ $\cong \begin{pmatrix} \pi^m & p^i \\ & y^2 \end{pmatrix}$. If $|y^2| > |\alpha|$, we show how to express $L = \mathfrak{O}x + \mathfrak{O}y_1$ with $|y_1{}^2| < |y^2|$. Write $|y^2| = |\pi|^{m+r}$, $r \geqq 0$: if r is even, pick β in $\mathfrak{u}$ with $\beta^2 \equiv -y^2/\pi^r x^2 \pmod{\pi}$ and set $y_1 = \beta\pi^{r/2}x + y$; if r is odd, pick β in $\mathfrak{u}$ with $\beta^2 \equiv -y^2/pp^*\pi^{r-1}x^2$ $(\mathrm{bod}\,\pi)$ and set $y_1 = \beta p^*\pi^{(r-1)/2}x + y$. The usual computations show $y_1{}^2 = \{\pi^{m+r+1}\}$, and we proceed by induction. q. e. d.

The following result should be compared with its quadratic-form analogue, equation (18.1) of [8]:

PROPOSITION 10.3. *Let L be a $\mathfrak{p}^i$-modular lattice of rank at least 3. Then $L = L_0 \oplus (\bigoplus_\lambda H_\lambda)$, where L_0 is a $\mathfrak{p}^i$-modular line or plane with $\mathfrak{n}L_0 = \mathfrak{n}L$, and each $H_\lambda \cong H(i)$.*

Proof. By scaling, assume $i = 0$ or 1.

1) *L normal.* Here $i = 0$, and by Proposition 4.4, b), we can write $L = L_1 \oplus L_2$, where $L_1 = \mathfrak{O}x_1 \oplus \mathfrak{O}x_2 \oplus \mathfrak{O}x_3$. Since $\operatorname{rank} L_1 = 3$, L_1 is isotropic, and so by Proposition 4.2, $L_1 = (\mathfrak{O}y_1 + \mathfrak{O}y_2) \oplus \mathfrak{O}y_3 \cong \begin{pmatrix} 0 & 1 \\ & [\pi^r] \end{pmatrix} \oplus ([1])$ with $r \geqq 0$. Pick α in $\mathfrak{u}$ with $\alpha^2 \equiv -y_2{}^2/(pp^*)^r y_3{}^2 \pmod{\pi}$, define $z_2 = y_2 + \alpha p^r y_3$, and note that $y_1 \cdot z_2 = 1$ and $|z_2{}^2| \leqq |\pi|^{r+1}$; thus $\mathfrak{O}y_1 + \mathfrak{O}z_2$

is a component of L_1, i. e., $L_1 \cong \begin{pmatrix} 0 & 1 \\ & \{\pi^{r+1}\} \end{pmatrix} \oplus ([1])$. Repeating this procedure, we may as well assume to begin with that $y_2{}^2\mathfrak{O} \subseteq nH(0)$, i. e., $n(\mathfrak{O}y_1 + \mathfrak{O}y_2) = nH(0)$; applying Proposition 9.2, $\mathfrak{O}y_1 + \mathfrak{O}y_2 \cong H(0)$. We conclude by induction on the rank of $(\mathfrak{O}y_1 + \mathfrak{O}y_2)^{\perp}$, where $n((\mathfrak{O}y_1 + \mathfrak{O}y_2)^{\perp}) = nL$ by (4.3).

2) *L subnormal.* Here $L = L_1 \oplus \cdots \oplus L_t$, with each L_λ a p^i-modular plane, and with, say, $nL_1 = nL = p^{2m}\mathfrak{O}$. Write

$$L_1 = \mathfrak{O}x_1 + \mathfrak{O}y_1 \cong \begin{pmatrix} [\pi^m] & p^i \\ & \{\pi^m\} \end{pmatrix}$$

and

$$L_2 = \mathfrak{O}x_2 + \mathfrak{O}y_2 \cong \begin{pmatrix} [\pi^{m+r}] & p^i \\ & [\pi^{m+s}] \end{pmatrix},$$

with $r, s \geqq 0$. Pick α, β in $\mathfrak{u}$ with $\alpha^2 \equiv -x_2{}^2/(pp^*)^r x_1{}^2 \pmod{\pi}$ and $\beta^2 \equiv -y_2{}^2/(pp^*)^s x_1{}^2 \pmod{\pi}$, set $x_3 = x_2 + \alpha p^r x_1$ and $y_3 = y_2 + \beta p^s x_1$, and note that $K_2 = \mathfrak{O}x_3 + \mathfrak{O}y_3 \cong \begin{pmatrix} \{\pi^{m+r+1}\} & [p^i] \\ & \{\pi^{m+s+1}\} \end{pmatrix}$; thus K_2 is a component of $L_1 \oplus L_2$, say $L_1 \oplus L_2 = K_1 \oplus K_2$ where necessarily $nK_1 = nL_1 = nL$. As in the normal case, we may take $x_2{}^2\mathfrak{O} \subseteq nH(i)$ and $y_2{}^2\mathfrak{O} \subseteq nH(i)$; thus $L_2 \cong H(i)$, and we conclude as above.

q. e. d.

We are now in a position to determine the invariants for modular lattices:

PROPOSITION 10.4. *Let L and K be p^i-modular lattices, i any ordinal. Then L and K are isometric if and only if rank L = rank K, nL = nK, and dL ≅ dK.*

Proof. We need prove only the sufficiency. Applying the cancellation law (Proposition 9.3) to the decomposition given in Proposition 10.3, it suffices to assume L and K both planes (the result for lines is clear by $dL \cong dK$). As a further reduction, we may (by scaling) take $i = 0$ or 1, and assume (by Proposition 9.2) that $nL = nK = p^{2m}\mathfrak{O} \supset nH(i)$ and (Proposition 10.2, a)) L and K are subnormal. Write

$$L = \mathfrak{O}x_1 + \mathfrak{O}x_2 \cong \begin{pmatrix} \pi^m & p^i \\ & \{\alpha\} \end{pmatrix} \text{ and } K = \mathfrak{O}y_1 + \mathfrak{O}y_2 \cong \begin{pmatrix} \pi^m & p^i \\ & \{\alpha\} \end{pmatrix},$$

with α as in Proposition 10.2. Let j be the ordinal determined by

$$p^j\mathfrak{O} = 2p^{-2m+2i}\mathfrak{O} \text{ in R-P, } p^j\mathfrak{O} = 2p^{-2k-2m-1+2i}\mathfrak{O} \text{ in R-U,}$$

and define the plane $H_0 = \mathfrak{O}x_3 + \mathfrak{O}x_4 \cong \begin{pmatrix} 0 & p^j \\ y_2{}^2 - x_2{}^2 \end{pmatrix}$; using (9.1), we see that $nH_0 = nH(j)$ in each case, hence $H_0 \cong H(j)$ by Proposition 9.2, a). The sublattice $L_1 = \mathfrak{O}x_1 + \mathfrak{O}(x_2 + x_4)$ of $L \oplus H_0$ is clearly isometric to K; noting further that $sL_1 = sL \supset sH_0$ (since $p^{2m}\mathfrak{O} \supset nH(i)$ by assumption), (4.7) shows that the p^i-modular L_1 is a component of $L \oplus H_0$, say $L \oplus H_0 = L_1 \oplus J$, with J a p^j-modular plane. We claim $nJ = nH_0$: using (4.15), $\mathfrak{R}L = p^{-i}L$ and $\mathfrak{R}H_0 = p^{-j}H_0$, so (by (9.1)) $n\mathfrak{R}H_0 = n\mathfrak{R}L$ in all cases; then

$$n\mathfrak{R}L_1 + n\mathfrak{R}J = n\mathfrak{R}\,(L_1 \oplus J) = n\mathfrak{R}\,(L \oplus H_0) = n\mathfrak{R}L + n\mathfrak{R}H_0 = n\mathfrak{R}H_0,$$

i.e., $n\mathfrak{R}J \subseteq n\mathfrak{R}H_0$, thus

$$nJ = n\mathfrak{R}\,\mathfrak{R}J = n(p^j\mathfrak{R}J) = p^{2j}n\mathfrak{R}J \subseteq p^{2j}n\mathfrak{R}H_0 = n(p^j\mathfrak{R}H_0) = nH_0,$$

and so $nJ = nH_0$ by Proposition 9.1, a). Also,

$$dJ \cong d(L_1 \oplus J)/dL_1 \cong d(L \oplus H_0)/dK \cong dH_0;$$

by Proposition 9.2, c), $J \cong H_0$. We may now cancel (Proposition 9.3) to obtain $L \cong L_1$; but $L_1 \cong K$. q. e. d.

11. The ramified dyadic case. The invariants. We consider a new concept. Let L, K be lattices with $|\,dL\,| = |\,dK\,|$ and with bases $\langle x_\lambda \rangle$ and $\langle y_\lambda \rangle$, respectively, such that $\det((x_\lambda \cdot x_\mu))/\det((y_\lambda \cdot y_\mu)) = 1 + \alpha$, where $|\,\alpha\,| < 1$. We shall write $dL/dK \cong 1 \pmod{\alpha\mathfrak{O}}$, and say that L and K have the same *determinant* $\mathit{mod}\,\alpha\mathfrak{O}$; this is an equivalence relation. Clearly, $dL_1/dK_1 \cong 1 \pmod{\alpha\mathfrak{O}}$ and $dL_2/dK_2 \cong 1 \pmod{\alpha\mathfrak{O}}$ implies

$$d(L_1 \oplus L_2)/d(K_1 \oplus K_2) \cong 1 \pmod{\alpha\mathfrak{O}};$$

furthermore, $dL \cong dK$ implies $dL/dK \cong 1 \pmod{\alpha\mathfrak{O}}$ for any α in $\mathfrak{p}$.

PROPOSITION 11.1. *Let* $\bigoplus_{1 \leqq \lambda \leqq t} L_\lambda$ *and* $\bigoplus_{1 \leqq \lambda \leqq t} K_\lambda$ *be two Jordan splittings for the same lattice* L. *Suppose* L_1 *(and therefore* K_1*) is either a* p^0*-modular line or a* p*-modular plane. Then* $dL_1/dK_1 \cong 1 \pmod{\mathfrak{f}(1)}$.

Proof. 1) *L_1 a p^0-modular line.* Write $L_1 = \mathfrak{O}x$ and $K_1 = \mathfrak{O}y$, with $|\,x^2\,| = |\,y^2\,| = 1$. Express $y = ax + z$ with a in $\mathfrak{O}$ and z in $\bigoplus_{2 \leqq \lambda \leqq t} L_\lambda$, thus $x^2/y^2 = 1/aa^* - z^2/aa^*y^2$; since $|\,x^2/y^2\,| = 1$ and $|\,z^2/y^2\,| = |\,z^2\,| \leqq |\,p\,|^{u(2)} = |\,p\,|^{u(1)+u(2)-2s(1)}$ (here $u(1) = s(1) = 0$), a must be a unit, thus z^2/aa^*y^2 is in $\mathfrak{f}(1)$, and $dL_1/dK_1 \cong 1 \pmod{\mathfrak{f}(1)}$.

2) *L_1 a p-modular plane.* Write $L_1 = \mathfrak{O}x_1 + \mathfrak{O}x_2$ and $K_1 = \mathfrak{O}y_1 + \mathfrak{O}y_2$ with $x_1 \cdot x_2 = y_1 \cdot y_2 = p$, and express $y_1 = X_1 + z_1$ and $y_2 = X_2 + z_2$, with

X_1, X_2 in L_1, and z_1, z_2 in $\bigoplus_{2\leqq\lambda\leqq t} L_\lambda$. Computing with (9.1), we find $[2\operatorname{Re}(y_\lambda \cdot z_\lambda)]\mathfrak{O} \subseteq nH(s(2)) \subseteq p^{u(2)}\mathfrak{O}$, $\lambda = 1, 2$, thus $X_\lambda^2 \equiv y_\lambda^2 \pmod{p^{u(2)}}$, therefore $\pmod{p^{s(2)}}$, and also $X_1 \cdot X_2 \equiv y_1 \cdot y_2 \pmod{p^{s(2)}}$; this shows $|d(\mathfrak{O}X_1 + \mathfrak{O}X_2)| = |dK_1| = |dL_1|$ (cf. (4.8)), and so $L_1 = \mathfrak{O}X_1 + \mathfrak{O}X_2$ by (2.6). We now wish to express $L_1 = \mathfrak{O}X + \mathfrak{O}X_2 \cong \begin{pmatrix} y_1^2 + \{p^{u(2)}\} & p \\ \cdot & y_2^2 + \{p^{u(2)}\} \end{pmatrix}$: from above, $(y_1 \cdot y_2)/(X_1 \cdot X_2) = 1 + p^{s(2)-1}a$ for some integer a, and we set $X = (1 + p^{s(2)-1}a)X_1$. Clearly, $X \cdot X_2 = y_1 \cdot y_2 = p$. To show $X^2 = y_1^2 + \{p^{u(2)}\}$, we must estimate $2\operatorname{Re}(p^{s(2)-1}a)X_1^2$ and $(p^{s(2)-1}aX_1)^2$: as with $2\operatorname{Re}(y_\lambda \cdot z_\lambda)$, above, we have $[2\operatorname{Re}(p^{s(2)-1}a)]\mathfrak{O} \subseteq nH(s(2)-1) = n[p^{-1}H(s(2))] \subseteq p^{u(2)-2}\mathfrak{O}$, hence $|2\operatorname{Re}(p^{s(2)-1}a)X_1^2| \leqq |p|^{u(2)}$; and since $p^{s(2)-1}aX_1$ lies in $p^{s(2)-1}L_1 = (L_1)_{(s(2))}$ (cf. (4.7)), it follows that $|(p^{s(2)-1}aX_1)^2| \leqq |p|^{u(2)}$. Thus $X^2 = X_1^2 + \{p^{u(2)}\} = y_1^2 + \{p^{u(2)}\}$, giving the desired expression. But then, using (4.9),

$$X^2X_2^2 - (X \cdot X_2)(X_2 \cdot X) \equiv y_1^2y_2^2 - (y_1 \cdot y_2)(y_2 \cdot y_1) \pmod{p^{u(1)+u(2)}},$$

so $dL_1/dK_1 \cong 1 \pmod{\mathfrak{f}(1)}$.

q. e. d.

PROPOSITION 11.2. *Let* $\bigoplus_{1\leqq\lambda\leqq t} L_\lambda$ *and* $\bigoplus_{1\leqq\lambda\leqq t} K_\lambda$ *be two Jordan splittings for a lattice* L *of the following special type: each* p^{2i}*-modular component is a line, each* p^{2i+1}*-modular component a plane. Fix* j, $1 \leqq j \leqq t$. *Then there exists a finite chain* $\bigoplus L_\lambda \to \cdots \to \bigoplus I_\lambda \to \bigoplus J_\lambda \to \cdots \to \bigoplus K_\lambda$ *of Jordan splittings such that, for any adjacent pair* $\bigoplus I_\lambda$, $\bigoplus J_\lambda$, *either*

(i) $\bigoplus_{1\leqq\lambda\leqq j} I_\lambda = \bigoplus_{1\leqq\lambda\leqq j} J_\lambda$, *or*

(ii) $I_\lambda = J_\lambda$ *for all* $\lambda \leqq j-1$.

A proof of this result can be easily based on that of the analogous Proposition 8 of [9], and will not be given here; in fact, our hypothesis on the type of L allows us to replace the "op" change of basis in [9], pp. 163-164, by suitable use of our Proposition 4.2—details left to the reader. Note that a "canonical basis" for L—mentioned in [9]—is merely one formed by taking the union of separate bases for each component in some Jordan splitting.

LEMMA. *Let* L *be a* p^i*-modular lattice of rank at least* 4, *and suppose* α *is any* E*-scalar in* nL. *Then* $L = L_1 \oplus L_2$, *where* L_1 *represents* α *and* $nL_2 = nL$.

Proof. By Propositions 10.3 and 9.1, we may assume $nL \supset nH(i)$.

If L is normal, then (Proposition 4.4) $L = L_1 \oplus L_2$, where L_1 is a plane and $nL_1 = nL_2 = nL$, and the result follows from Proposition 10.1. Suppose then L is subnormal, and write $L = L_0 \oplus H$, where $nL_0 = nL$ and $H = \mathfrak{O}x_1 + \mathfrak{O}x_2 \cong \begin{pmatrix} 0 & p^i \\ & 0 \end{pmatrix} \cong H(i)$. By Proposition 10.1, L_0 contains a vector y with $y^2 = \alpha$, and so the plane $L_1 = \mathfrak{O}(x_1 + y) + \mathfrak{O}x_2$ represents α and is a component of L: $L = L_1 \oplus L_2$. Now, $nL_1 = \alpha\mathfrak{O} + nH(i)$, i.e., $\alpha\mathfrak{O} \subseteq nL$. If $\alpha\mathfrak{O} \subset nL$, then $nL_1 \subset nL$, so $nL_2 = nL$; if $\alpha\mathfrak{O} = nL$, then we note by inspection that $z = y - \alpha(x_2 \cdot x_1)^{-1}x_2$ lies in $L_1{}^{\perp} = L_2$ and satisfies $z^2 = \alpha$, so here too $nL_2 = \alpha\mathfrak{O} = nL$. q. e. d.

PROPOSITION 11.3. *Let $L = L_1 \oplus L_2$ be a saturated Jordan splitting, with L_1 a p^i-modular line or plane, $i = 0$ or 1, and with rank $L_2 \geqq 4$. Let K_1 be a p^i-modular lattice for which rank $K_1 =$ rank L_1, $nK_1 = p^{u(1)}\mathfrak{O}$, and $dL_1/dK_1 \cong 1$ $(\mathrm{mod}\, \mathfrak{f}(1))$. Then L has a second saturated Jordan splitting $L_1' \oplus L_2'$ with $L_1' \cong K_1$.*

Proof. We first shall express L_1 and K_1 as follows: $L_1 = \mathfrak{O}x_1$ and $K_1 = \mathfrak{O}y_1$ if L_1 is a line, $L_1 = \mathfrak{O}x_1 + \mathfrak{O}x_2$ and $K_1 = \mathfrak{O}y_1 + \mathfrak{O}y_2$ if L_1 is a plane, where $x_1{}^2 \equiv y_1{}^2$ $(\mathrm{mod}\, p^{u(2)})$, and (if a plane) $x_1 \cdot x_2 = y_1 \cdot y_2 = p^{s(1)}$ and $x_2{}^2 = y_2{}^2$. If L_1 is a line, any x_1 and y_1 will work, since $\mathfrak{f}(1) = p^{u(2)}\mathfrak{O}$ here. In the case of a plane, use Proposition 10.2 to write $L_1 = \mathfrak{O}x_1 + \mathfrak{O}x_2$ $\cong \begin{pmatrix} \pi^{\frac{1}{2}u(1)} & p^{s(1)} \\ & \{\pi^{\frac{1}{2}u(1)}\} \end{pmatrix}$; then since $dL_1/dK_1 \cong 1$ $(\mathrm{mod}\, \mathfrak{f}(1))$, K_1 has a basis $\langle v_1, v_2 \rangle$ with $x_1{}^2 x_2{}^2 - (x_1 \cdot x_2)(x_2 \cdot x_1) = v_1{}^2 v_2{}^2 - (v_1 \cdot v_2)(v_2 \cdot v_1) + \alpha$, where $|\alpha| \leqq |p|^{u(1)+u(2)}$. The plane $J \cong \begin{pmatrix} x_2{}^2 - (\alpha/x_1{}^2) & p^{s(1)} \\ & x_1{}^2 \end{pmatrix}$ is evidently $p^{s(1)}$-modular, with $nJ = nK_1$ and $dJ \cong dK_1$, so by Proposition 10.4 is isometric to K_1; thus K_1 obviously contains the desired vectors y_1, y_2.

Now write $y_1{}^2 = x_1{}^2 + \beta$, with $|\beta| \leqq |p|^{u(2)}$. Since $L_1 \oplus L_2$ is saturated and rank $L_2 \geqq 4$, the Lemma shows $L_2 = J_1 \oplus J_2$, where J_1 contains a vector z with $z^2 = \beta$, and $nJ_2 = p^{u(2)}\mathfrak{O}$. Define $L_1' = \mathfrak{O}(x_1 + z)$ if L_1 is a line, $L_1' = \mathfrak{O}(x_1 + z) + \mathfrak{O}x_2$ if a plane. L_1' is a sublattice of $L_1 \oplus J_1$, and is isometric to K_1; in particular, by (4.7), L_1' is a component of $L_1 \oplus J_1$, so $L = L_1' \oplus L_2'$. Finally, $L_1' \cong K_1$ implies $L_1' \oplus L_2'$ is saturated at $s(1)$, while $J_2 \subseteq L_2' \subseteq L_{(s(2))}$ implies $p^{u(2)}\mathfrak{O} = nJ_2 \subseteq nL_2' \subseteq n(L_{(s(2))}) = p^{u(2)}\mathfrak{O}$, which is saturation at $s(2)$. q. e. d.

THEOREM 11.4. *Suppose F/E is ramified dyadic. Let L and K be lattices, and suppose $L = \bigoplus_{1 \leqq \lambda \leqq t} L_\lambda$ and $K = \bigoplus_{1 \leqq \lambda \leqq T} K_\lambda$ are any Jordan splittings. Then L and K are isometric if and only if the following four conditions hold:*

1) *L and K are of the same type.*

2) $dL \cong dK$.

3) $u_L(\lambda) = u_K(\lambda)$ *for all* $\lambda \leqq t$.

4) $d(L_1 \oplus \cdots \oplus L_j)/d(K_1 \oplus \cdots \oplus K_j) \cong 1$ (*mad* $\mathfrak{f}(j)$)
for all $j \leqq t-1$.

Proof. Necessity. 1), 2), and 3) have already been shown. To show 4), we may assume each p^{2i}-modular component is a line, each p^{2i+1}-modular component a plane, since adjunction of a (π^i) or an $H(i)$ to both L and K, or cancellation of an $H(i)$ from both L and K, will alter neither the isometry $L \cong K$ nor the determinant-ratio 4); and we adjoin and cancel in view of Proposition 10. 3, so as to retain as Jordan components of L and K only the L_0 of that Proposition. We may take $L = K$. By scaling (recall (4.13)), we may also assume $s(j) = 0$ or 1. Now apply Proposition 11. 2: since determinant mod $\alpha\mathfrak{O}$ is an equivalence relation, it suffices to prove 4) for splittings $\bigoplus L_\lambda$ and $\bigoplus K_\lambda$ that differ by a transformation of "type (ii)" (it is trivial for "type (i)"); thus say $L = \bigoplus_{1 \leqq \lambda \leqq t} L_\lambda = (\bigoplus_{1 \leqq \lambda \leqq j-1} L_\lambda) \oplus (\bigoplus_{j \leqq \lambda \leqq t} K_\lambda)$. But $\bigoplus_{j \leqq \lambda \leqq t} L_\lambda = (\bigoplus_{1 \leqq \lambda \leqq j-1} L_\lambda)^\perp = \bigoplus_{j \leqq \lambda \leqq t} K_\lambda$, and since the ideal $\mathfrak{f}(j)$ for $\bigoplus_{j \leqq \lambda \leqq t} L_\lambda$ is included in $\mathfrak{f}_L(j)$, we may take $j = 1$. But the result is now that of Proposition 11. 1.

Sufficiency. We use induction on the non-negative ordinal $s(t) - s(1)$: for $s(t) - s(1) = 0$, $L \cong K$ by Proposition 10. 4, so the induction can be started. There will be four preliminary reductions:

a) $s(1) = 0$ or 1 (by scaling).

b) $\operatorname{rank} L_1 = \operatorname{rank} K_1 \leqq 2$ (write L_1 and K_1 as in Proposition 10. 3; 1)-4) remain valid when the H_λ are cancelled, and after the smaller lattices are proved isometric, the H_λ may be restored).

c) $\operatorname{rank} L_2 = \operatorname{rank} K_2 \geqq 4$ (as in b), adjunction and cancellation of hyperbolic planes does not affect 1)-4) or the isometry).

d) $\bigoplus L_\lambda$ and $\bigoplus K_\lambda$ are saturated splittings (the necessity-proof shows that 1)-4) hold for *any* splittings of L and K).

Now use Proposition 11. 3 to write $L_1 \oplus L_2 = L_1' \oplus L_2'$ with $nL_2' = nL_2 = p^{u(2)}\mathfrak{O}$ and $L_1' \cong K_1$. The lattices $L' = L_2' \oplus L_3 \oplus \cdots \oplus L_t$ and $K' = \bigoplus_{2 \leqq \lambda \leqq t} K_\lambda$ clearly satisfy 1), 2), and 4); by saturation, $u_{L'}(\lambda) = u_L(\lambda)$

$= u_K(\lambda) = u_{K'}(\lambda)$ for all $\lambda \geqq 2$, so L' and K' also satisfy 3). By induction, $L' \cong K'$, hence $L \cong K$. q. e. d.

NOTE. Although we have found simpler criteria (Theorems 6.2, 7.1, 8.2) for isometry in the other cases, let us show that Theorem 11.4 is also correct there, thus giving a complete set of invariants in *all* cases (in the non-commutative case, we ignore the determinant conditions 2) and 4) in Theorem 11.4; in the ramified-non-dyadic and unramified cases, note that "determinant mod $\alpha\mathfrak{O}$" may be defined exactly as above, with the same properties). For the necessity, 1)-3) are again obvious; as for 4), we have the even stronger statement $d(L_1 \oplus \cdots \oplus L_j) \cong d(K_1 \oplus \cdots \oplus K_j)$ for all $j \leqq t-1$ (in the ramified non-dyadic case, cf. Proposition 8.1, b), and Theorem 8.2). For the sufficiency, condition 1) is enough to guarantee isometry in the non-commutative and unramified cases; in the ramified non-dyadic case, (4.12) and Hensel's lemma transform condition 4) into

$$\text{“} d(L_1 \oplus \cdots \oplus L_j) \cong d(K_1 \oplus \cdots \oplus K_j) \text{ for all } j \leqq t-1, \text{”}$$

which together with 2) implies $dL_\lambda \cong dK_\lambda$ for *all* λ—and then $L \cong K$ by Theorem 8.2.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] E. Artin, ***Algebraic Numbers and Algebraic Functions I***, Princeton University, New York University, 1951.

[2] ———, ***Geometric Algebra***, New York, 1957.

[3] N. Bourbaki, "Formes Sesquilineaires et Formes Quadratiques," ***Algebre, Chapitre IX***, Paris, 1959.

[4] M. Deuring, ***Algebren***, Berlin, 1935.

[5] W. H. Durfee, "Congruence of quadratic forms over valuation rings," ***Duke Mathematical Journal***, vol. 11 (1944), pp. 687-697.

[6] N. Jacobson, "A note on Hermitian forms," ***Bulletin of the American Mathematical Society***, vol. 46 (1940), pp. 264-268.

[7] B. W. Jones, ***The Arithmetic Theory of Quadratic Forms***, Mathematical Association of America, 1950.

[8] O. T. O'Meara, "Quadratic forms over local fields," ***American Journal of Mathematics***, vol. 77 (1955), pp. 87-116.

[9] ———, "Integral equivalence of quadratic forms in ramified local fields," ***ibid.***, vol. 79 (1957), pp. 157-186.

[10] ———, ***Introduction to Quadratic Forms***, to appear.

[11] O. F. G. Schilling, ***The Theory of Valuations***, New York, 1950.

[12] B. L. van der Waerden, ***Algebra, Zweiter Teil***, Vierte Auflage, Berlin, 1959.

COVARIANT FIBERINGS OF KLEIN SPACES, II.*

By G. D. Mostow.[1]

---

1. Introduction.

1. Introduction. By a *Klein* space, we mean a space on which a Lie group of transformations operates transitively, or equivalently, a quotient space of cosets G/G' where G is a Lie group and G' a closed subgroup. It is well-known that if G is a connected Lie group, then G is topologically the direct product of a compact subgroup and a euclidean subspace. From this, one can deduce in the case of a connected closed subgroup G' that G/G' admits a strong deformation retraction to a subspace K/K', where K is a maximal compact subgroup of G and $K' = K \cap G$ is a maximal compact subgroup of G'. This result suggests that the Klein space G/G' can indeed be decomposed into a compact Klein space K/K' and a euclidean subspace. In his paper entitled "On covariant fiberings of Klein spaces" hereafter referred to as CFI (cf. [3]) the author investigated this suggestion. An example of H. Samelson had shown that G/G' is *not* the direct product of K/K' and a euclidean space. CFI introduced the notion of a *covariant fibering* of a Klein space G/G': it is a fibering into euclidean fibers satisfing

1) the fibers are permuted transitively by the action of some maximal compact subgroup K of G.

2) the subgroup of K which keeps a fiber F invariant operates linearly on F.

It is proved in CFI that any two covariant fiberings are equivalent. Moreover the *existence* of covariant fiberings is established in the case that both G and G' are *connected* and G' is *self-adjoint modulo the radical* of G. (Cf. Section 6 below for definitions.) This last condition is satisfied in a wide variety of cases, notably when G' is semi-simple, compact, or in the radical of G.

Thus it was conjectured that the self-adjoint hypothesis could be dropped. The proof of this conjecture is precisely the purpose of this paper, and it settles in the affirmative Problem 38 of the Cornell Conference of 1955 on Fiber Bundles (W. Massey, "Some problems in algebraic topology and the theory of fiber bundles," *Annals of Mathematics*, vol. 62 (1955), pp. 327-359).

---

* Received January 26, 1962.

[1] Supported in part under National Science Foundation Grant 19575.

We rely on the results of CFI, and our method applies equally well to the case of a Klein space G/G', where G and G' are Lie groups having a *finite* number of connected components; such groups are referred to hereafter as *fcc* groups. The reader is referred to CFI for a more detailed account and bibliography of the results that we shall be citing below.

2. Let G be a fcc group. By a *decomposition* of G we shall mean a topological direct product decomposition

$$G = K \cdot E$$

where K is a compact subgroup of G and E is a closed subspace homeomorphic to euclidean space; that is, the map $(k, e) \rightarrow ke$ of $K \times E$ into G is a homeomorphism onto G. The subgroup K is necessarily a maximal compact subgroup of G. For if L is a compact subgroup of G with $L \supset K$, then the fibering of the coset space $K \diagdown G$ with fiber $K \diagdown L$ and base $K \diagdown L \diagdown K \diagdown G = L \diagdown G$ yields a fibering of the euclidean space E with compact fiber $K \diagdown L$; hence $K \diagdown L$ is a point and $L = K$.

Every fcc group admits a decomposition (cf. [4]) and a decomposition with factor K is tantamount to a cross section of the euclidean base $K \diagdown G$ into the principal K-bundle G. Moreover, the maximal compact subgroups of an fcc group are conjugate under an inner automorphism (cf. ibid.). It follows exactly as in the proof of Theorem 2. 3 in CFI that if K' and L' are compact subgroups of an fcc group G, and if K and L are maximal compact subgroups with $K \supset K'$ and $L \supset L'$, then any inner automorphism which sends K' onto L' can be modified to an inner automorphism having the same effect on K' and mapping K onto L.

3. Let G be an fcc group and G' a fcc subgroup. Let $K' \cdot E'$ be a decomposition of G'. By a *split extension* of $K' \cdot E'$, we shall mean a decomposition $G = K \cdot E$ such that

1) $K \supset K'$

2) $E = F \cdot E'$, with F a euclidean subspace of G; that is, the map $(f, e) \rightarrow fe$ of $F \times E'$ to E is a homeomorphism.

3) $kFk^{-1} = F$ for all $k \in K'$.

We shall say that G and G' have *coherent* decompositions if there exists a decomposition of G which is a split extension of a decomposition of G'.

Let $G \supset G' \supset G''$ be fcc groups with decompositions $K \cdot E$, $K' \cdot E'$, and $K'' \cdot E''$ respectively. If $K \cdot E$ is a split extension of $K' \cdot E'$, and if $K' \cdot E'$ is a split extension $K'' \cdot E''$, then it follows readily that G' and G'' are closed in G and G' respectively and that $K \cdot E$ is a split extension of $K'' \cdot E''$.

4. Let G be an fcc group with decompositions $K \cdot E$ and let G' be an fcc subgroup with decompositions $K' \cdot E'$ and $K' \cdot E_1'$. If $K \cdot E$ is a split extension of $K' \cdot E'$, then $E = F \cdot E'$ with $kFk^{-1} = F$ for all $k \in K'$. Set $E_1 = F \cdot E_1'$. We assert that $K \cdot E_1$ is a split extension of $K' \cdot E_1'$. For consider the mapping $t_1 \colon K \times F \times E_1' \to G$ given by $(k, f, e_1) \to kfe_1$. In order to prove that t_1 is a homeomorphism onto G, we consider the continuous map

$$u \colon K \times F \times E_1' \to K \times F \times E$$

defined by

$$(k, f, e_1) \to (kh, h^{-1}fh, e)$$

where $e_1 = he$ with $h \in K'$ and $e \in E'$. From the fact that $K' \cdot E'$ and $K' \cdot E_1'$ are decompositions of G', it follows at once that u^{-1} exists and is continuous. Clearly the map t_1 factors

$$t_1 = t \cdot u$$

where t denotes the homeomorphism $(k, f, e) \to kfe$ of $K \times F \times E'$ onto G. Consequently t_1 is a homeomorphism and $K \cdot E_1$ is a split extension of $K' \cdot E'$.

5. Let $G \supset G' \supset G''$ be fcc groups. If G and G' have coherent decompositions, and if G' and G'' have coherent decompositions, then G and G'' have coherent decompositions. For by hypothesis, we have decompositions $K \cdot E = G$, $K' \cdot E' = G' = K_1' \cdot E_1'$, $K'' \cdot E'' = G''$ with $K \cdot E$ and $K_1' \cdot E_1'$ split extensions of $K' \cdot E'$ and $K'' \cdot E''$ respectively. Inasmuch as the maximal compact subgroups of an fcc group are conjugate under an inner automorphism (cf. [4]) we can find an element $g \in G'$ with $gK'g^{-1} = K_1'$. Upon replacing the decompositions $K' \cdot E'$ and $K \cdot E$ by their transforms under the inner automorphism by the element g, we can assume that $K_1' = K'$. By paragraph 4, we can find a split extension $K \cdot E_1$ of $K' \cdot E_1'$. Then by paragraph 3, $K \cdot E_1$ is a split extension of $K'' \cdot E''$. Consequently, G *and* G'' have coherent decompositions.

6. We shall recall now the basic theorem from CFI on the existence of coherent decompositions. Let $\boldsymbol{G}$ be a semi-simple real Lie algebra. A subalgebra $\boldsymbol{K}$ is called *compact* if the analytic subgroup K of the adjoint group of $\boldsymbol{G}$ whose Lie algebra is $\operatorname{ad} \boldsymbol{K}$ is compact—or equivalently, if the Killing form $\operatorname{Tr} \operatorname{ad} X \operatorname{ad} Y$ is negative definite on $\boldsymbol{K}$ and K is closed. A *Cartan decomposition* of $\boldsymbol{G}$ is a decomposition $\boldsymbol{G} = \boldsymbol{K} + \boldsymbol{E}$ where $\boldsymbol{K}$ is a maximal

compact subalgebra and $\boldsymbol{E}$ is the orthogonal complement to $\boldsymbol{K}$ with respect to the Killing form. It is known that all the maximal compact subalgebras are conjugate by an automorphism of the adjoint group; and that the linear map of G which sends $X \rightarrow X$ for all $X \in \boldsymbol{K}$ and $Y \rightarrow -Y$ for all $Y \in \boldsymbol{E}$ is an automorphism of the Lie algebra $\boldsymbol{G}$, called the ***K**-star automorphism.* A subalgebra of G is called *self-adjoint* if it is stable under some star-automorphism. This condition is equivalent to the condition that the subalgebra have a self-adjoint image under some faithful representation.

A sufficient condition that a subalgebra $\boldsymbol{H}$ of the semi-simple Lie algebra $\boldsymbol{G}$ be self-adjoint is that the subalgebra $\mathrm{ad}\,\boldsymbol{H}$ of $\mathrm{ad}\,\boldsymbol{G}$ be algebraic and fully reducible (cf. [4]); a necessary and sufficient condition is that $\rho(\boldsymbol{H})$ be a self-adjoint family of matrices for some finite dimensional faithful representation ρ of $\boldsymbol{G}$.

Let G be a simply connected semi-simple analytic group, let $\boldsymbol{K}$ be a maximal compact subalgebra of the Lie algebra $\boldsymbol{G}$, and let K be the analytic group corresponding to $\boldsymbol{K}$. Since $\mathrm{Ad}\,G/\mathrm{Ad}\,K$ is euclidean and simply connected, we find that $K = \mathrm{Ad}^{-1}(\mathrm{Ad}\,K)$, and thus K contains the center of G. It follows at once that the center of G is elementwise fixed under the automorphism of G induced by the $\boldsymbol{K}$-star automorphism of $\boldsymbol{G}$ and thus the $\boldsymbol{K}$-star-automorphism of $\boldsymbol{G}$ induces a well-defined *K-star-automorphism* of *any* analytic group with Lie algebra $\boldsymbol{G}$; moreover this automorphism leaves fixed each element in the center of G.

We can now define a subgroup H of an analytic semi-simple group G to be *self-adjoint in G* if H is stable under some star-automorphism of G.

If H is connected, then H is self-adjoint in G if and only if $\boldsymbol{H}$ is self-adjoint in $\boldsymbol{G}$. If H is not connected, the following sufficient condition covers our needs: a closed fcc subgroup H of an analytic semi-simple group G is self-adjoint in G if its Lie algebra $\boldsymbol{H}$ is self-adjoint in $\boldsymbol{G}$ and has an algebraic image under the adjoint representation (cf. [4]).

If G is an analytic group with radical R, and if H is a subgroup of G, we call H *self-adjoint modulo the radical of G* if HR/R is self-adjoint in G/R.

Theorem 2.1 of CFI can be restated as:

THEOREM. *Let G be an analytic group and let G' be a closed analytic subgroup. If G' is self-adjoint in G modulo the radical, then G' and G have coherent decompositions.*

7. The main result of this paper involves dropping the hypothesis that G' is *self-adjoint* in G modulo the radical. At the same time, we wish to

treat the somewhat more general case of *fcc* groups rather than ***analytic*** groups.

Let G be an fcc group and let G_1 denote the connected components of its identity. Then G and G_1 have coherent decompositions. For if $K \cdot E$ is any decomposition of G, and if $K_1 = K \cap G_1$ then $K_1 \cdot E$ is a decomposition of G_1 and $K \cdot E$ is a split extension of $K_1 \cdot E$. Thus, by paragraph 5, in proving for any pair G, G' of fcc groups with G' a closed subgroup of G, that G and G' have coherent decompositions, no generality is lost in assuming that G is an analytic group.

An examination of the proof of the above-cited Theorem 2.1 of CFI reveals that one can prove in exactly the same way:

THEOREM. *Let G be an analytic group, and let G' be a closed fcc subgroup. If G' is self-adjoint in G modulo the radical, then G' and G have coherent decompositions.*

Indeed, in the proof of the cited theorem, the only assertion which requires amplification occurs in Case 3, the case that G is semi-simple. It is asserted there that if C is self-adjoint in G with respect to the K-star automorphism of G, then

$$C = C \cap K \cdot E'$$

where $\boldsymbol{K} + \boldsymbol{E}$ is a Cartan decomposition of $\boldsymbol{G}$ and $E' = \exp(\boldsymbol{C} \cap \boldsymbol{E})$. However this fact follows at once from [6], Section 3, where it is proved for the case of fcc matric groups G.

8. A closed subgroup G' of a Lie group G is called *almost maximal* in G if $\dim G' < \dim G$ and G' is not contained in any closed subgroup H with $\dim G' < \dim \mathrm{H} < \dim G$. Clearly every closed subgroup G' of a Lie group G lies in some almost maximal subgroup if $\dim G' < \dim G$. Also, every closed fcc subgroup G' lies in some almost maximal fcc subgroup of G if $\dim G' < \dim G$.

We now prove a result which is the analogue of Theorem 3.1 in [5].

LEMMA. *Let G be an analytic semi-simple group, and let M be an almost maximal subgroup. Let $\boldsymbol{G}$ and $\boldsymbol{M}$ denote the Lie algebras of G and M respectively. Then the image of $\boldsymbol{M}$ under the adjoint representation of $\boldsymbol{G}$ is an algebraic Lie algebra and either*

i) *$\boldsymbol{M}$ is self-adjoint in $\boldsymbol{G}$.*

or

ii) *$\boldsymbol{G} = \boldsymbol{K} + \boldsymbol{M}$ for any maximal compact subalgebra $\mathbf{K}$, and $G = KM$, where $\mathbf{K}$ is the Lie algebra of the analytic subgroup K.*

Proof. Let M_1 denote the connected component of the identity in M. If M_1 is normal in G, then $\mathbf{M}$ is an ideal in $\mathbf{G}$, and therefore $\mathbf{M}$ is self-adjoint in $\mathbf{G}$ and ad $\mathbf{M}$ is an algebraic Lie algebra.

We can assume therefore that M_1 is not normal in G. Since M_1 is normal in M, no generality is lost in assuming that M is the normalizer of M_1. Then M contains the center of G. We can therefore transfer the discussion to the adjoint group of G; thereby G may be taken as the adjoint group of $\mathbf{G}$ and M becomes the normalizer of $\mathbf{M}$ in G—an fcc group of finite index in its algebraic group hull. Thus $\mathbf{M}$ is an algebraic Lie subalgebra of $\mathbf{G}$.

Let $\mathbf{N}$ be the maximum ideal of nilpotent elements in the radical of $\mathbf{M}$. If $\mathbf{N}=0$, then $\mathbf{M}$ is fully reducible (cf. [2]) and therefore self-adjoint (cf. [4]). Thus it remains only to consider the case that $\mathbf{N}\neq 0$.

Since $\mathbf{N}$ is invariant under the inner automorphisms by the elements of $\mathbf{M}$, the normalizer of $\mathbf{N}$, which is a proper subgroup of G, cannot exceed M in dimension. Thus $\mathbf{M}$ is the Lie algebra normalizer of $\mathbf{N}$ in $\mathbf{G}$.

It follows by the argument of [5] p. 509 that $\mathbf{M}$ is the orthogonal subspace to $\mathbf{N}$ with respect to the Killing form of $\mathbf{G}$ and that $\mathbf{M}$ contains a maximal triangular subalgebra of $\mathbf{G}$. Consequently (cf. ibid.)

$$\mathbf{G}=\mathbf{K}+\mathbf{M} \text{ and } G=KM$$

where $\mathbf{K}$ is some maximal compact subalgebra of $\mathbf{G}$. Inasmuch as all the maximal compact subalgebras are conjugate under an inner automorphism from G, and since G is generated by K and M, it follows that all compact subalgebras are conjugate under an inner automorphism from M, and thus $G=KM$ and $\mathbf{G}=\mathbf{K}+\mathbf{M}$ for all maximal compact subalgebras $\mathbf{K}$.

9. Let G be an analytic group, and let H be an almost maximal closed fcc subgroup. Let H_1 denote the connected component of the identity in H. Let R denote the radical of G. If H does not include R, then $\overline{H_1R}=G$. Dividing out the radical R, we find in the semi-simple analytic group G/R, that

$$G/R=[G/R,G/R]=[\overline{H_1R}/R,\overline{H_1R}/R]=[H_1R/R,H_1R/R]\subset HR/R,$$

where the Poisson brackets denote the commutator subgroup (the last equality is by the well-known result of A. Malcev that the commutator subgroup of an analytic subgroup J coincides with the commutator subgroup of its topological closure $\bar{J}$; it is proved by transferring the question to the simply connected covering group of $\bar{J}$, in which every *normal* analytic subgroup is closed). Thus $G = HR$ and H is self-adjoint modulo the radical. G and H have coherent decompositions by paragraph 7 in this case.

If $H \supset R$, then H/R is an almost maximal closed fcc subgroup of the semi-simple analytic group G/R. Set $M = H/R$. By the Lemma of paragraph 8, the image of $\mathbf{M}$ under the adjoint representation of $\mathbf{G/R}$ is an analytic Lie algebra and either

i) $\mathbf{M}$ is self-adjoint in $\mathbf{G/R}$, or

ii) $\mathbf{G/R} = \mathbf{J^*} + \mathbf{M}$, where $\mathbf{J^*}$ is any maximal compact subalgebra of $\mathbf{G/R}$.

In case i), the fcc subgroup M is self-adjoint in G/R by the remark of paragraph 6. Hence H is self-adjoint in G modulo the radical, and in this case too, G and H have coherent decompositions.

In case ii), let ρ denote the adjoint representation of G. Let $\mathbf{G} = \mathbf{S} + \mathbf{R}$ be a Levi decomposition of the Lie algebra $\mathbf{G}$, and let S denote the semi-simple analytic group corresponding to $\mathbf{S}$. Then $\rho(H \cap S)$ is an almost maximal subgroup of S and

$$\rho(S) = J^* \rho(H \cap S)$$

for any maximal compact subgroup J^* of the semi-simple linear group $\rho(S)$. Consequently

$$\rho(G) = J \rho(H)$$

where J is a maximal compact subgroup of $\rho(G)$, and

$$G = LH$$

where $L = \rho^{-1}(J)$. We have

$$H/H \cap L = G/L = \rho(G)/J.$$

Since $\rho(G)/J$ is homeomorphic to euclidean space, and since every bundle with euclidean base is a product bundle, it follows that there is a euclidean subspace E^0 in H such that

$$H = (H \cap L) \cdot E^0 \qquad \text{(direct)}$$

and

$$G = L \cdot E^0 \qquad \text{(direct)}$$

Inasmuch as $\rho(L)$ is compact, we know that L is the direct group product of its maximum compact subgroup K and a central vector subgroup. It follows at once that there are vector central subgroups V_1 and V_2 in L and $H \cap L$ respectively such that

$$L = (K(H \cap L)) \cdot V_1 \qquad \text{(direct)}$$
$$H \cap L = (H \cap K) \cdot V_2 \qquad \text{(direct)}$$

and

$$K(H \cap L) = K \cdot V_2 \qquad \text{(direct)}.$$

Consequently

$$G = L \cdot E^0 = K \cdot V_2 \cdot V_1 \cdot E^0 = K \cdot V_1 \cdot V_2 \cdot E^0 \quad \text{(direct)}$$

Setting $F = V_1$ and $E' = V_2E^0$ we have

$$G = KFE'$$

and

$$H = (H \cap K)E'$$

with $kFk^{-1} = F$ for all $k \in H \cap K$. Thus G and H have coherent compositions.

10. Let G be an fcc group and let G' be a closed fcc subgroup, not of the same dimension. Since G' lies in an almost maximal closed fcc subgroup of G, we can find an ascending chain of closed fcc subgroups.

$$G' = G^{(1)} \subset G^{(2)} \subset \cdots \subset G^{(n)} = G$$

with $G^{(i)}$ almost maximal in $G^{(i+1)}$ for each $i = 1, \cdots, n-1$. By the results of paragraphs 9 and 7, each pair of $G^{(i)}$ and $G^{(i+1)}$ have coherent decompositions. Consequently G' and G have coherent decompositions by paragraph 5. On the other hand, if $\dim G' = \dim G$, then G and G' hace coherent decompositions; namely, if $K' \cdot E$ is a decomposition of G', *then* $G = K \cdot E$ is a split extension of $K' \cdot E$ for any therefore proved that G and G' have coherent decompositions.

Taking into account the remark in paragraph 1 on the conjugacy of pairs of compact subgroups $K \supset K'$ with K maximal compact, we can restate our result as follows.

THEOREM A. *Let G be a Lie group having a finite number of connected components, and let G' be a closed subgroup having a finite number of connected components. Let K' be a maximal compact subgroup of G' and let K be any maximal compact subgroup of G with $K \supset K'$. Then there exists closed euclidean subspaces F and E' in G such that*

$$G' = K' \cdot E'$$

and

$$G = K \cdot F \cdot E'$$

as topological direct products, and

$$kFk^{-1} = F$$

for all $k \in K'$.

From Theorem A it follows at once that

$$G/G' = K \times_{K'} F.$$

In this way we have proved

THEOREM. *Let G be an fcc group and G′ an fcc closed subgroup. Then G/G′ admits a covariant fibering.*

YALE UNIVERSITY.

REFERENCES.

[1] C. Chevalley, *Theorie des groupes de Lie*, tome 2, Hermann, Paris, 1951.
[2] ———, tome 3, Hermann, Paris, 1956.
[3] G. D. Mostow, "On covariant fiberings of Klein spaces," *American Journal of Mathematics*, vol. 77 (1955), pp. 247-277.
[4] ———, "Self-adjoint groups," *Annals of Mathematics*, vol. 62 (1955), pp. 44-55.
[5] ———, "on maximal subgroups of real Lie groups," *Annals of Mathematics*, vol. 74 (1961), pp. 503-517.
[6] ———, "Homogeneous spaces of finite invariant measure," *Annals of Mathematics*, vol. 75 (1962), pp. 17-37.

SOME PROPERTIES OF GENERALIZED AXIALLY SYMMETRIC POTENTIALS.*[1]

By R. P. GILBERT.

I. Introduction.

The theory of integral operators as developed by S. Bergman, and some others (see [1]-[8]), formalized a procedure by which analytic functions of one or several complex variables are transformed into solutions of partial differential equations. The importance of this theory is three-fold (i) to transplant theorems concerning analytic functions into theorems about solutions of partial differential equations, (ii) to obtain representation theorems for solutions, and (iii) to consider the analytic continuation of solutions.

Recently, S. Bergman [4], [5] has found certain properties of harmonic functions in three variables which depend on the series development of the function. In this note we shall show that several of Professor Bergman's results have analogies in the case of solutions of the partial differential equation

$$L[u] \equiv \frac{\partial^2 u}{\partial^2 u} + \frac{\partial z^2}{\partial \rho^2} + \frac{k}{\rho}\frac{\partial u}{\partial \rho} = 0, \tag{1}$$

where $k \geqq 0$. The study of solution of this equation has been referred to by Professor Weinstein [9], [10] as generalized axially symmetric potential theory (GASPT). This equation is initially arrived at by considering those solutions of the n-dimensional Laplace-equation, which depend solely on $z = x_1$, $\rho = (x_2^2 + \cdots + x_n^2)^{\frac{1}{2}}$. In this case k is the integer, $n - 2$. The location of the sigularities of solutions of (1) have been investigated by Szegö [11], Nehari [12], Erdélyi [13], Henrici [20], and Gilbert [14].

II. GASPT with meromorphic associates.

In previous works [14], [15] this author has introduced an integral operator (closely related to the Sonine representation), which transforms analytic functions of one complex variable into GASPT function elements $\Phi(p)$. More specifically,

$$u(z,\rho) \equiv \Phi(p) = A_k(f, L, p^0) \equiv \int_L f(\sigma)\,(\zeta - \zeta^{-1})^{k-1} \frac{d\zeta}{\zeta}, \tag{2}$$

---

* Received February 9, 1962.

[1] This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49(638)228.

where $\sigma = z + \frac{i\rho}{2}(\zeta + \zeta^{-1})$, $|p - p^0| < \epsilon$, $p \equiv z + i\rho$, $p^0 \equiv z^0 + i\rho^0$, $\epsilon > 0$ is sufficiently small, and the integration path L is the upper semi-circular arc connecting $+1$ to -1. ($f(\sigma)$ is called the associate of the GASPT $\Phi(p)$.)

In order to see how this mapping occurs, we consider the case where $f(\sigma)$ is analytic in a neighborhood of $\sigma = 0$, that is for $|\sigma|$ sufficiently small. Then it may be seen that $A_k(f, L, 0)$ defines the GASPT function element [14]

$$\Phi(p) = i(4i)^{k-1}\Gamma(k/2)^2 \sum_{n=0}^{\infty} \frac{a_n n!}{\Gamma(k+n)} r^n C_n^{k/2}(\cos\theta), \tag{4}$$

where $r^2 = z^2 + \rho^2$, $\cos(\theta) = z/r$, and $C_n^{\lambda}(\cos\theta)$ is a Gegenbauer polynomial.

We next consider the case where $f(\sigma)$ is meromorphic in the finite σ-plane. In this case we may use the Mittag-Leffler theorem [16], as does Bergman [5], to obtain a representation for the transformed functions. For instance, one may write

$$f(\sigma) = \sum_{\nu=1}^{\infty} \left[P_\nu\left(\frac{1}{\sigma - b_\nu}\right) - p_\nu(\sigma)\right] + e(\sigma), \tag{5}$$

where the b_ν and the poles of $f(\sigma)$, $P_\nu\left(\frac{1}{\sigma - b_\nu}\right)$ the corresponding principal parts, the $p_\nu(\sigma)$ suitably chosen polynomials to ensure convergence, and $e(\sigma)$ an entire function. Moreover, since this series converges uniformly in every compact set of the σ-plane which does not contain the b_ν, one may evaluate the integral representation for $\Phi(p)$ by inverting the order of integration and summation.

$$\begin{aligned} \Phi(p) &= A_k(f, L, p^0) \equiv \int_L f(\sigma)(\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta \\ &= \left\{\sum_{\nu=1}^{\infty}\left[P_\nu\left(\frac{1}{\sigma - b_\nu}\right) - p_\nu(\sigma)\right] + e(\sigma)\right\}(\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta \\ &= \sum_{\nu=1}^{\infty}\left[\Phi_\nu(p) - \phi_\nu(p)\right] + E(p), \end{aligned} \tag{6}$$

where $\phi_\nu(p)$ is a GASPT polynomial, and $E(p)$ is an entire GASPT. The integrals,

$$\Phi_\nu(p) = \int_L P_\nu\left(\frac{1}{\sigma - b_\nu}\right)(\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta, \tag{7}$$

are easily evaluated when k is an odd integer; in this case one may replace L

by the unit circle, divide by 2, and apply the residue theorem. For instance, let

$$P_\nu\left(\frac{1}{\sigma-b_\nu}\right)=\sum_{\mu=1}^{m_\nu}\frac{M_{\mu\nu}}{(\sigma-b_\nu)^\mu},$$

then

$$\begin{aligned}\Phi(p)&=\sum_{\mu=1}^{m_\nu}M_{\mu\nu}\int_L\frac{(\zeta-\zeta^{-1})^{k-1}}{(\sigma-b_\nu)^\mu}\frac{d\zeta}{\zeta}\\ &=\pi i\sum_{\mu=1}^{m_\nu}M_{\mu\nu}\frac{(-1)^\mu}{\mu!}\left(\frac{2}{i\rho}\right)^k\frac{\partial^\mu}{\partial z^\mu}[(z-b_\nu)^2+\rho^2]^{k/2}.\end{aligned}\tag{8}$$

Hence, if k is an odd integer, $\Phi(p)$ has the representation

$$\begin{aligned}\Phi(p)=\pi i\sum_{\nu=0}^{\infty}\{[\sum_{\mu=1}^{m_\nu}M_{\mu\nu}\frac{(-1)^\mu}{\mu!}\left(\frac{2}{i\rho}\right)^k\frac{\partial^\mu}{\partial z^\mu}([z-b_\nu]^2+\rho^2)^{k/2}]-\phi_\nu(z,\rho)\}\\ +E(z,\rho).\end{aligned}\tag{9}$$

One may show, as Bergman has done in the case of harmonic functions in three variables, that this infinite series representation converges in every compact set in the p-plane, which does not contain singularities of the $\Phi_\nu(p)$. It has been shown earlier [14] for $k>0$, that if $f(\sigma)$ has a singularity at $\sigma=b$, then the GASPT has a singularity at $p=b$, and $\bar{b}$. For the case where $f(\sigma)$ is meromorphic, and k odd, this is evident from the representations (8) and (9).

One may show the series converges by proceeding along the lines in which the Mittag-Leffler theorem is proved [5], [16]. In order to show that the series $\sum_{\nu=1}^{\infty}[\Phi_\nu(p)-\phi_\nu(p)]$ converges, we first recall that the polynomials $p_\nu(\sigma)$ were chosen such that $\left|P_\nu\left(\frac{1}{\sigma-b_\nu}\right)-p_\nu(\sigma)\right|<\delta_\nu$, for $|\sigma|\leqq|b_\nu|\lambda$, $\lambda<1$, and that $\sum_{\nu=1}^{\infty}\delta_\nu<\infty$. Now, if the integration path L is restricted so that $\zeta=e^{i\psi}$, $0\leqq\psi\leqq\pi$, one has $|\sigma|=|z+i\rho\cos\psi|\leqq|z+i\rho|=|p|$, and hence one has $|\sigma|\leqq R$, when $|p|\leqq R$. One may show that if $R\neq|b_\nu|$, the the series $\sum_{\nu=N+1}^{\infty}[\Phi_\nu(p)-\phi_\nu(p)]$ converges uniformly and absolutely in $|p|\leqq R\lambda$, if N is chosen such that $|b_\nu|<R$ for $\nu\leqq N$, and $|b_\nu|>R$ for $\nu>N$. This may be seen as follows:

$$\begin{aligned} \left| \sum_{\nu=N+1}^{\infty} [\Phi_\nu(p) - \phi_\nu(p)] \right| &\leqq \sum_{\nu=N+1}^{\infty} | \Phi_\nu(p) - \phi_\nu(p) | \\ &\leqq \sum_{\nu=N+1}^{\infty} \left| \int_L [P_\nu \left(\frac{1}{\sigma - b_\nu} \right) - p_\nu(\sigma)] (\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta \right| \\ &\leqq 2^{k-1} \sum_{\nu=N+1}^{\infty} \int_L \left| P_\nu \left(\frac{1}{\sigma - b_\nu} \right) - p_\nu(\sigma) \right| \cdot | d\zeta/\zeta . = 2^{k-1}\pi \sum_{\nu=1}^{\infty} \delta_\nu < \infty. \end{aligned} \tag{10}$$

The first N terms of the original series have pole-like, branch points in $| p | \leqq \mathrm{R}$; by deleting closed neighborhoods of these points from $| p | \leqq R$ the series is seen to converge uniformly and absolutely in this closed set.

As Bergman [4] has shown in the case of harmonic functions in three variables, one may obtain an interesting class of GASP with meromorphic associates by considering

$$f(\sigma) = \frac{\Gamma'(\sigma + \alpha)}{\Gamma(\sigma + \alpha)} = -\gamma + \sum_{n=0}^{\infty} \frac{1}{n+1} - \frac{1}{n + \sigma + \alpha}, \quad \alpha \neq 0 \tag{11}$$

where γ is Euler's constant and $\Gamma(\sigma)$ is the gamma-function. Bergman [4] has also shown, that since there are numerous relations involving $\Gamma'(\sigma)/\Gamma(\sigma)$ one may transplant these into relations concerning the class of harmonic functions. We shall list these results for the class of GASP $\{\Phi_k(p;\alpha)\}$ defined by

$$\Phi_k(p;\alpha) = \int_L \frac{\Gamma'(\sigma - \alpha)}{\Gamma(\sigma - \alpha)} (\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta, \quad \alpha \neq 0. \tag{12}$$

First, one notes that $\Phi_k(p;\alpha)$ has pole-like branch points at $p = \alpha - n$, and $\bar{\alpha} - n$. From the relations [4]

$$\frac{\Gamma'(\sigma)}{\Gamma(\sigma)} = \log \sigma - 1/2\sigma - \tfrac{1}{2} \sum_{\mu=3}^{\infty} (\mu - 3) \mu^{-1} \sum_{\mu=1}^{\infty} (\sigma - \nu)^{-\mu}, \quad \mathrm{Re}\, \sigma > 0, \tag{13}$$

and

$$\frac{\Gamma'(m\sigma)}{\Gamma(m\sigma)} = 1/m \sum_{\nu=0}^{m-1} \frac{\Gamma'(\sigma + \nu/m)}{\Gamma(\sigma + \nu/m)} + \log m, \qquad m = 2, 3, 4, \cdots, \tag{14}$$

one obtains respectively,

$$\Phi_k(p;\alpha) = \mathscr{L}_k(p;\alpha) - \tfrac{1}{2}\mathscr{F}_k^{(1)}(p;\alpha) - \tfrac{1}{2} \sum_{\mu=3}^{\infty} (\mu - 2) \mu^{-1} \sum_{\nu=1}^{\infty} \mathscr{F}_k^{(\mu)}(p + \nu;\alpha),$$

where

$$\mathscr{L}_k(p;\alpha) = \int_L \log(\sigma + \alpha) (\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta, \tag{15}$$

$$\mathscr{F}_k^{(\mu)}(p + \nu;\alpha) = \int_L (\sigma + \nu + \alpha)^{-\mu} (\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta,$$

for $\operatorname{Re} p > 0$, and

$$(16) \qquad \Phi_k(mp;m\alpha) = 1/m \sum_{\nu=0}^{m-1} \Phi_k(p+\nu/m;\alpha) + \log m \int_L (\zeta - \zeta^{-1})^{k-1} d\zeta/\zeta.$$

III. Singularities of GASPT. The use of integral operators to study the solutions of linear partial differential equations has been particularly profitable in the investigation of their singularities. The reason for this is, that frequently theorems may be stated in terms of the solutions themselves, and not the associated analytic functions. In the case of GASPT there is a one-to-one mapping of singularities of the analytic function $f(\sigma)$ onto the singularities of $\Phi(p)$, for $I_m\, p \geqq 0$. Because of this, one can readily transplant theorems concerning singularities of analytic functions into theorems concerning GASPT [14], [15]. For instance, a Hadamard multiplication of singularities theorem [17], [18] also holds for GASPT.

THEOREM 1. *Let* $\Phi(p)$, *and* $\Psi(p)$ *be two* GASP *with the series expansions*

$$\Phi(p) = \sum_{n=0}^{\infty} a_n r^n C_n^{k/2}(\cos\theta),$$

$$\Psi(p) = \sum_{n=0}^{\infty} b_n r^n C_n^{k/2}(\cos\theta),$$

Furthermore, let us suppose, that $\Phi(p)$ *and* $\Psi(p)$ *have singuarities respectively at the point-pairs* $\{\alpha. \bar{\alpha}\}$, *and* $\{\beta, \bar{\beta}\}$. *Then, the GASP given by the development*

$$F(p) = \sum_{n=0}^{\infty} a_n b_n r^n C_n^{k/2}(\cos\theta),$$

has singularities at either the point pair $\{\alpha\beta, \overline{\alpha\beta}\}$ *or at* $\{\bar{\alpha}\beta, \alpha\bar{\beta}\}$.

Proof. Suppose $f(\sigma)$, $g(\sigma)$, $h(\sigma)$ are the associates of $\Phi(p)$, $\Psi(p)$, $F(p)$, respectively. By the original Hadamard theorem, we know that if $f(\sigma)$, $g(\sigma)$ have singularities at δ, γ respectively, $h(\sigma)$ may be singular only at $\delta\gamma$. The corresponding singularities for $\Phi(p)$, $\Psi(p)$, and $F(p)$ are then at $\{\delta, \bar{\delta}\}$, $\{\gamma, \bar{\gamma}\}$, and $\{\delta\gamma, \overline{\delta\gamma}\}$ respectively. This concludes our proof.

In a similar way, one may transplant theorems about regions of regularity. For instance, we may obtain the radius of convergence given by the usual Hadamard formula involving the expansion coefficients.

THEOREM 2. *Let* $\Phi(p)$ *be a* GASP *with the following series development about the origin.*

$$\Phi(p) = \sum_{n=0}^{\infty} a_n r^n C_n^{k/2}(\cos\theta).$$

Then $\Phi(p)$ *converges uniformly and absolutely in any compact sub set of the disk of convergence,*

$$|p| < R \text{ where } R^{-1} = \overline{\lim_{n\to\infty}}\, |a_n|^{1/n}.$$

First, we note that the associated function

$$f(\sigma) = \frac{-i}{(4i)^{k-1}}\,\Gamma(k/2)^{-2}\sum_{n=0}^{\infty}\frac{\Gamma(k+n)}{n!}\,a_n\sigma^n$$

has the radius of convergence

$$\begin{aligned} R &= [\overline{\lim_{n\to\infty}}\,|\frac{\Gamma(k+n)}{n!}\,a_n|^{1/n}]^{-1} \\ (17)\qquad &= (\overline{\lim_{n\to\infty}}\,[\frac{\Gamma(k+n)}{n!}]^{1/n}\,\overline{\lim_{n\to\infty}}\,.\,a_n|^{1/n})^{-1} \\ &= (\overline{\lim_{n\to\infty}}\,|a_n|^{1/n})^{-1}. \end{aligned}$$

This may be seen by using the Stirling formula expansion of $\frac{\Gamma(k+n)}{n!}$ for large n, which shows that if $k>1$, $\frac{\Gamma(k+n)}{n!}$ tends to ∞ as n^{k-1}, and if $0<k\leqq 1$, $\frac{\Gamma(k+n)}{n!}$ tends to 0 as $n^{k-1}e^{k-1}$.

Next, we recall that $|\sigma| = |z+i\rho\cos\psi| \leqq |p|$, when $\zeta = e^{i\psi}$, hence if one chooses p such that $|p| \leqq R_1 < R_2 < R$ then $|\sigma| \leqq R_1$ also. Moreover, since the series expansion for $f(\sigma)$ converges uniformly and absolutely in $|\sigma| \leqq R_2$ one has

$$\begin{aligned} (18)\qquad |\sum_{n=0}^{\infty} a_n r^n C_n^{k/2}(\cos\theta)| &\leqq \sum_{n=0}^{\infty} |a_n r^n C_n^{k/2}(\cos\theta)| \\ &\leqq \sum_{n=0}^{\infty}\int_0^{\pi} |b_n|\cdot|\sigma|^n\cdot|2\sin\psi|^{k-1}d\psi \leqq \frac{M}{1-R_2/R}, \end{aligned}$$

where $b_n = \frac{-ia_n\Gamma(k+n)}{(4i)^{k-1}n!}$. This establishes our result.

In general, over convergence and gap theorems may also be transplanted to the case of GASPT. We list as an example the analogous result of a theorem due to Fabry [19].

THEOREM 3. *If* $\Phi(p) = \sum_{n=0}^{\infty} a_n r^n C_n^{k/2}(\cos\theta)$ *contains infinitely many suffixes* m_i *such that*

(19) (i) $$\lim_{i\to\infty} |a_{m_i}|^{1/m_i} = 1/R,$$

and

(ii) $$\lim_{i\to\infty} s_i/m_i = 0,$$

where s_i is the number of non-vanishing coefficients in the expansion for $\Phi(p)$, which satisfy the inequality

$$m_i(1-\theta) < n < m_i(1+\theta),\ n \neq m_i,\ 0<\theta<1.$$

Then $\Phi(p)$ has as a natural boundary the circle $|p| = R$.

IV. Inequalities for GASPT. In general theorems concerning inequalities for GASPT are given in terms of the analytic associate. Occasionally it is possible by means of the inverse operator [14] A_k^{-1}:

$$f(\sigma) = \frac{k}{(2i)^k\pi}\int_{-1}^{+1} W(r,\xi)\,\frac{(1-\xi^2)^{k/2}(1-(\sigma^2/r^2))\,d\xi}{(1-2\xi(\sigma/r)+(\sigma^2/r^2))^{k/2+1}}, \tag{20}$$

(where $W(r,\xi) = u(r, \cos^{-1}\xi) \equiv \Phi(p)$, and C is the real axis) to obtain the inequality in terms of GASPT alone. For the general case, $k>0$, we list the following *weak* Schwarz-lemma.

THEOREM 4. *Let $\Phi(p)$ be a* GASP *with an associate $f(\sigma)$, which has the following properties:*

(i) $$|f(\sigma)| \leqq 1, \text{ when } |\sigma| \leqq 1,$$

(ii) $$f(0) = 0,$$

then $|\Phi(p)| \leqq \pi^k/k\,|p|$, when $|p| \leqq 1$.

Since $|p| \leqq 1$ implies $|\sigma| \leqq 1$ (for $\zeta = e^{i\psi}$), this result follows from the inequality

$$|\Phi(p)| \leqq \int_0^{\pi} |f(\sigma)|\cdot|2\sin\psi|^{k-1}\,d\psi \leqq 2^k \int_0^{\pi/2} \psi^{k-1}\,d\psi = \pi^k/k. \tag{21}$$

We next consider bounds for GASPT with entire associate functions. From Theorem 2 we realize, that if $\Phi(p) = \sum_{n=0}^{\infty} a_n r^n C_n^{k/2}(\cos\theta)$ has an entire associate, then $\overline{\lim_{n\to\infty}} |a_n|^{1/n} \to 0$. If $M(r)$ is the maximum modulus of $f(\sigma)$ on $|\sigma| = r$, then the order of $f(\sigma)$ is defined to be [19],

$$\lambda = \overline{\lim_{r\to\infty}} \frac{\log\log M(r)}{\log r}.$$

In other words λ is the lower limit of all $\mu \geqq 0$, such that $M(r) \leqq e^{r^\mu}$ for r large enough. Suppose the order of $f(\sigma)$ is λ, but that one may replace the inequality $M(r) \leqq e^{r^\lambda}$ by $M(r) \leqq e^{ar^\lambda}$, $a > 0$. We now may further distinguish our functions by considering the lower limit a_L of values $\{a\}$ for which the inequality is true. If $a_L = 0$, $f(\sigma)$ is said to be of *minimum type*, if $a_L > 0$, $f(\sigma)$ is of *normal type*, and if no $a_L > 0$ satisfies the inequality $f(\sigma)$ is of *maximum type* [19].

By the same sort of argument used in Equation (21) one realizes that if the associate function $f(\sigma)$ satisfies the inequality $|f(\sigma)| \leqq e^{ar^\lambda}$ for $|\sigma| = r$ sufficiently large, then $|\Phi(p)| \leqq e^{ar^\lambda}$ for $|p| = r$ sufficiently large. This however, does not indicate $\Phi(p)$ is of order λ. Suppose $f(\sigma)$ is given as in Equation (3), and $\Phi(p)$ is given by (4), then at $\cos\theta = 1$ $(r = z,\ \rho = 0)$ one has

$$(22)\qquad \begin{aligned} \Phi(z) &= \frac{i(4i)^{k-1}}{\Gamma(k/2)^2} \sum_{n=0}^{\infty} a_n \frac{n!}{\Gamma(k+n)} z^n C_n^{k/2}(1) \\ &= \frac{i(4i)^{k-1}}{\Gamma(k/2)^2 \Gamma(k)} \sum_{n=0}^{\infty} a_n z^n. \end{aligned}$$

Now, if the maximum modulus of $f(\sigma)$ occurs, for $|\sigma|$ sufficiently large, along the real axis it would be clear that $\Phi(z)$ and $f(\sigma)$ were of the same order and type.[2]

One may also obtain bounds for GASPT with meromorphic associates (see Bergman [5]). Suppose $f(\sigma) = h(\sigma)/g(\sigma)$, where $h(\sigma)$ and $g(\sigma)$ are entire functions of order λ, then we may obtain bounds for $|\Phi(p)|$ in certain regions of the p-plane by applying the theory of entire functions to $h(\sigma)$, $g(\sigma)$. For instance, if $m(r)$ is the minimum modulus of $g(\sigma)$, and $\{\sigma_n\}$ is the set of zeros of $g(\sigma)$, then $m(r) > e^{-r^{\lambda+\epsilon}}$ on circles of arbitrarily large radius for those regions $\mathfrak{R}$ excluding the circles $|\sigma - \sigma_n| \leqq |\sigma_n|^{-h}$, where $h > \lambda$, the order of $g(\sigma)$. This result has an interpretation for GASPT. If no $\sigma_n = 0$, then $\Phi(p)$ is regular at $p = 0$. The domain of association of $\Phi(p)$ is the p-plane less the segments $E\{z = \operatorname{Re}\sigma_n, \rho^2 \geqq \operatorname{Im}\sigma_n; n = 1, 2, 3 \cdots\}$. Following Bergman, we then consider that subset of the p-plane less strips of thickness $2|\sigma_n|^{-h}$ covering the above segments, that is we consider the domain given by

$$\begin{aligned} \mathfrak{B} &\equiv E\{\sum_{\nu=1}^{\infty} [\operatorname{Re}\sigma_\nu + |\sigma_\nu|^{-h} \leqq z \leqq \operatorname{Re}\sigma_{\nu+1} - |\sigma_{\nu+1}|^{-h}, \rho^2 < \infty]\} \\ \cup\, &E\{\sum_{\nu=1}^{\infty} [\operatorname{Re}\sigma_\nu - |\sigma_\nu|^{-h} \leqq z \leqq \operatorname{Re}\sigma_\nu + |\sigma_\nu|^{-h}, \rho^2 < |\,|\sigma_\nu|^{-2h} - |z - \sigma_\nu|^2\,|]\}. \end{aligned}$$

---

[2] It is a conjecture of the author that a *modified* Schwarz-lemma holds for *GASPT*, and he is presently engaged in investigating this problem.

It is clear that if $p \in \mathfrak{B}$ then $\sigma \in \mathfrak{R}$, hence we have the analogy to Bergman's theorem, for the case of GASPT.

THEOREM 5. *Let* $\Phi(p)$ *be a* GASPT *with the associate* $f(\sigma) = h(\sigma)/g(\sigma)$, *where* $h(\sigma)$ *and* $g(\sigma)$ *are entire functions of order* λ. *Furthermore, let* $h(\sigma)$ *not be of maximum type. Then in the subdomain* $\mathfrak{B}$ *of the domain of association for* $\Phi(p)$, *and for* $|p| = r$ *sufficiently large one has the inequality*

$$|\Phi(p)| \leqq \pi 2^{k-1} e^{a' r^{\lambda+\epsilon'}},$$

where a' *and* ϵ' *are constants,* $\epsilon' > 0$ *is arbitrarily small.*

From the integral representation for $\Phi(p)$ it follows that,

$$|\Phi(p)| \leqq \int_L \left|\frac{h(\sigma)}{g(\sigma)}\right| \cdot |\zeta - \zeta^{-1}| \left|\frac{d\zeta}{\zeta}\right| \leqq 2^{k-1} \int_0^\pi \left|\frac{h(z+i\rho\cos\psi)}{g(z+i\rho\cos\psi)}\right| d\psi$$

$$\leqq \frac{2^{k-1}\pi e^{(a+\epsilon)r^\lambda}}{e^{-r^{\lambda+\epsilon'}}} = \pi 2^{k-1} e^{r^{\lambda+\epsilon'}[(a+\epsilon)r^{-\epsilon'}+1]},$$

where $a < \infty$ since $h(\sigma)$ is not of maximum type, and where $\epsilon > 0, \epsilon' > 0$ are arbitrarily small. If r is large enough (and $r > 1$) we then have

$$|\Phi(p)| \leqq \pi 2^{k-1} e^{a' r^{\lambda+\epsilon'}},$$

where $a' = a + 1 + \epsilon$.

THE INSTITUTE FOR FLUID DYNAMICS AND APPLIED MATHEMATICS,
UNIVERSITY OF MARYLAND.

REFERENCES.

[1] S. Bergman, "Zur Theorie der ein- und mehrwertigen harmonischen Funktionen des dreidimensional Raumes," *Mathematische Zeitschrift*, vol. 24 (1926), pp. 641-669.

[2] ———, "Zur Theorie der algebraischen Potential Funktionen des dreidimensional Raumes," *Mathematische Annalen*, vol. 99 (1928), pp. 629-659, and vol. 101 (1909), pp. 534-538.

[3] ———, "Properties of solutions of certain differential equations in three variables," *Journal of Mathematics and Mechanics*, vol. 7 (1958), pp. 805-820.

[4] ———, "Essential singularities of solutions of a class of linear partial differential equations in three variables," *Journal of Rational Mechanics and Analysis*, vol. 3 (1954), No. 5, pp. 539-560.

[5] ———, "Some properties of a Harmonic Function of three variables given by its series development," *Archive for Rational Mechanics and Analysis*, vol. 8 (1961), No. 3, pp. 207-222.

[6] R. Gilbert, "Singularities of three-dimensional harmonic functions," *Pacific Journal of Mathematics*, vol. 10 (1961), pp. 1243-1255.

[7] E. Kreyszig, "On regular and singular harmonic functions of three variables," *Archive for Rational Mechanics and Analysis*, vol. 4 (1960), pp. 352-370.

[8] ———, "Bergman operators for the three and four dimensional Laplace equations," *Archive for Rational Mechanics and Analysis* (to appear).

[9] A. Weinstein, "Discontinuous integrals and generalized potential theory," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 342-354.

[10] ———, "Generalized axially symmetric potential theory," *Bulletin of the American Mathematical Society*, vol. 59 (1953), pp. 20-38.

[11] G. Szegö, "On the singularities of zonal harmonic expansions," *Journal of Rational Mechanics and Analysis*, vol. 3 (1954), pp. 561-564.

[12] Z. Nehari, "On the singuarities of Legendre expansions," *Journal of Rational Mechanics and Analysis*, vol. 5 (1956), pp. 987-992.

[13] A. Erdélyi, "Singularities of generalized axially symmetric potentials," *Communications on Pure and Applied Mathematics*, vol. 9 (1956), pp. 403-414.

[14] R. Gilbert, "On the singularities of generalized axially symmetric potentials," vol. 6 (1960), No. 2, pp. 171-176.

[15] ———, "On generalized axially symmetric potentials," *Technical Note* BN-262, AFOSR-1613, October (1961).

[16] L. Ahlfors, *Complex Analysis*, McGraw-Hill (1953).

[17] E. C. Titchmarsh, *The Theory of Functions*, Oxford (1939).

[18] J. Hadamard, "Theoreme sur les series entieres," *Acta Mathematica*, vol. 22 (1898), pp. 55-64.

[19] P. Dienes, *The Taylor Series*, Oxford (1931).

[20] P. Henrici, "On the domain of regularity of generalized axially symmetric potentials," *Proceedings of the American Mathematical Society*, vol. 8 (1957), No. 1, pp. 29-31.

SOME NUMERICAL CRITERIA FOR CONTRACTABILITY OF CURVES ON ALGEBRAIC SURFACES.*[1]

By MICHAEL ARTIN.

Section 0.

Let V be a complete algebraic surface over an algebraically closed field k and let X be a connected curve on V. We assume that V is nonsingular along X. Call X *contractable* if there exists a map $\pi\colon V \to \bar{V}$ onto a surface $\bar{V}$ which is an isomorphism at every point of $V - X$ and such that $\pi(X)$ is a single point P. If X is contractable then π is determined uniquely by the condition that P be a normal point of $\bar{V}$. We call this uniquely determined map π the *contraction* of X.

An example of Hironaka shows that in general there are no numerical criteria equivalent with contractability of a given curve X. On the other hand, Castelnuovo's characterization $p(X) = 0$, $(X^2) = -1$ of irreducible exceptional curves of the first kind is numerical. It seems that if these conditions are read 'X sufficiently rational and negative' one obtains perhaps the strongest available numerical criterion for contractability. This (cf. (2.3)) and some more special results are in Section 2. The question of what is meant by a rational cycle is treated in Theorem (1.7).

I would like to thank H. Hironaka for some stimulating discussions on the subject and J. Tate for suggesting a simplification in the proof of Lemma (1.4).

Section 1.

Let $X = \cup X_i$ $(i = 1, \cdots, s)$ be a closed curve on a complete algebraic surface V/k, the X_i being its irreducible components. *We assume throughout the paper that V is nonsingular at every point of X.* By a positive cycle $Z = \sum r_i X_i$ $(r_i > 0)$ we mean the closed subscheme [2] of V defined by the coherent sheaf of ideals on V whose sections on an open $U \subset V$ are the rational functions $f \in \Gamma(U, \mathcal{O}_V)$ which have zeros of order at least r_i along X_i for all i with $X_i \cap U \neq \emptyset$. The local ring $\mathcal{O}_{Z,P}$ of a point P on Z is therefore $\mathcal{O}_{V,P}/(f)$ where $f = 0$ is a local equation for $\sum r_i X_i$ at P. We recall the

---

* Received February 20, 1962.

[1] The work on this paper was supported (in part) by the Army Research Office (Durham).

definition of the arithmetic genus $p(Z)=\frac{1}{2}((Z^2)+(Z\cdot K))+1$. (Since V is nonsingular along X the terms are defined.) $p(Z)$ satisfies the relation

$$(1.1)\qquad 1-p(Z)=\chi(Z)=\dim H^0(\mathcal{O}_Z)-\dim H^1(\mathcal{O}_Z).$$

If D is any divisional cycle on V we denote by $\mathrm{Tr}_Z D$ the restriction to Z of the invertible sheaf $\mathcal{L}(D)$, i. e., the cokernel of the canonical injection $\mathcal{L}(D-Z)\rightarrow\mathcal{L}(D)$ (these sheaves are defined at least in a neighborhood of Z).

Denote by $\mathcal{O}_Z{}^*$ the sheaf of units of $\mathcal{O}_Z$. It is well known that the elements of $H^1(\mathcal{O}_Z{}^*)$ correspond canonically to isomorphism classes of invertible sheaves on Z. Using the natural map $\epsilon_i\colon X_i\rightarrow Z$ we can associate with each invertible sheaf $\mathcal{L}$ on Z the sheaf $\epsilon_i{}^*(\mathcal{L})$ on X_i. Set $d_i(\mathcal{L})=$ degree $\epsilon_i{}^*(\mathcal{L})$ for $i=1,\cdots,s$ and $d(\mathcal{L})=(d_1(\mathcal{L}),\cdots,d_s(\mathcal{L}))\in\mathbf{Z}^s$. We call $d(\mathcal{L})$ the *degree* of $\mathcal{L}$. The map $\mathcal{L}\rightarrow d(\mathcal{L})$ yields a homomorphism (which we denote also by d)

$$(1.2)\qquad d\colon H^1(\mathcal{O}_Z{}^*)\rightarrow\mathbf{Z}^s.$$

It is easily seen that d is surjective.

We wish to get some information about $H^1(\mathcal{O}_Z{}^*)$ which will be needed later. Fix a positive connected cycle $Z=\sum r_iX_i$ $(i=1,\cdots,s)$. Then $X=\sum X_i=Z_{red}$ is the reduced scheme associated to Z and we have a surjective map $\mathcal{O}_Z\rightarrow\mathcal{O}_X$ of sheaves. Therefore the induced map $\mathcal{O}_Z{}^*\rightarrow\mathcal{O}_X{}^*$ is also surjective. Let N, M respectively denote the kernels of these maps. Since Z (and hence X) is connected, $H^0(\mathcal{O}_X)\approx k$ and $H^0(\mathcal{O}_X{}^*)\approx k^*$. Therefore the maps $H^0(\mathcal{O}_Z)\rightarrow H^0(\mathcal{O}_X)$ and $H^0(\mathcal{O}_Z{}^*)\rightarrow H^0(\mathcal{O}_X{}^*)$ are surjective. Taking into account the fact that X has dimension 1 we find exact sequences

$$(1.3)\qquad \begin{array}{l} 0\rightarrow H^1(N)\rightarrow H^1(\mathcal{O}_Z)\rightarrow H^1(\mathcal{O}_X)\rightarrow 0\\ 0\rightarrow H^1(M)\rightarrow H^1(\mathcal{O}_Z{}^*)\rightarrow H^1(\mathcal{O}_X{}^*)\rightarrow 0.\end{array}$$

The kernels $H^1(N)$ and $H^1(M)$ are related as follows:

LEMMA (1.4). *There exist chains of subgroups*

$$H^1(N)=N_0\supseteq N_1\supseteq\cdots\supseteq N_n=0,\ H^1(M)=M_0\supseteq M_1\supseteq\cdots\supseteq M_n=0$$

and isomorphisms $N_{j-1}/N_j\approx M_{j-1}/M_j$ $(j=1,\cdots,n)$ *where* $n=\sum(r_i-1)$.

Proof. We use induction on $n=\sum(r_i-1)$. If $n=0$, i. e., Z is reduced, the assertion is trivial. Assume therefore that the lemma is true for $n-1$

$(n>0)$. Choose an i with $r_i>1$ and set $Z'=Z-X_i$. Let N', M', J, K be the kernels of the maps $\mathcal{O}_{Z'}\to\mathcal{O}_X$, $\mathcal{O}_{Z'}^*\to\mathcal{O}_X^*$, $\mathcal{O}_Z\to\mathcal{O}_{Z'}$, $\mathcal{O}_Z^*\to\mathcal{O}_{Z'}^*$ respectively. Here $J^2=0$, in fact $JN=0$, and hence there is an isomorphism $e\colon J\to K$ given by $x\to 1+x$ for any section x of the sheaf J. Moreover there are exact sequences $0\to J\to N\to N'\to 0$, $0\to K\to M\to M'\to 0$. We obtain exact cohomology sequences

$$\begin{gathered} H^0(N')\to H^1(J)\to H^1(N)\to H^1(N')\to 0 \\ H^0(M')\to H^1(K)\to H^1(M)\to H^1(M')\to 0. \end{gathered} \tag{1.5}$$

Therefore the theorem will follow from the induction assumption applied to Z', if we can show that in the sequences (1.5) the images of $H^1(J)$, $H^1(K)$ in $H^1(N)$, $H^1(M)$, respectively, are isomorphic. To show this it suffices to show that the image of $H^0(N')$ in $H^1(J)$ is carried onto the image of $H^0(M')$ in $H^1(K)$ by the isomorphism $\epsilon\colon H^1(J)\to H^1(K)$ induced by e. Now there is a bijective map $\epsilon'\colon H^0(N')\to H^0(M')$ given by $s'\to 1+s'$. This map is in general not a group isomorphism. However we claim that the diagram

$$\begin{tikzcd} H^0(N') \arrow[r, "\delta"] \arrow[d, "\epsilon'"] & H^1(J) \arrow[d, "\epsilon"] \\ H^0(M') \arrow[r, "\delta"] & H^1(K) \end{tikzcd}$$

of maps commutes, where the horizontal arrows are those in (1.5). This will complete the proof.

Let $s'\in H^0(N')$ be given and choose an open covering $\{U_i\}$ of X such that s' can be lifted to sections s_i of N on U_i. Then $s_i-s_j\in\Gamma(U_i\cap U_j, J)$ and $\delta(s')$ is represented by the 1-cocycle $\{(s_i-s_j)\}$. Hence $\epsilon\delta(s')$ is represented by the cocycle $\{1+s_i-s_j\}$. On the other hand the section $\epsilon'(s')=1+s'$ can be lifted to the sections $1+s_i$ of M on U_i and therefore $\delta\epsilon'(s')$ is represented by the cocycle $\{(1+s_i)/(1+s_j)\}$. But since $s_i-s_j\in\Gamma(U_i\cap U_j, J)$ and $JN=0$ we have $(s_i-s_j)s_j=0$. Hence

$$(1+s_j)(1+s_i-s_j)=1+s_i,\ \text{i.e.,}\ (1+s_i)/(1+s_j)=1+s_i-s_j.$$

Thus the two cocycles are equal and the proof is complete.

Consider now the reduced curve $X=\sum X_i$. Denote by $\tilde{X}=\coprod X_i$ the direct sum of the curves X_i (cf. [2]). The map $\pi\colon\tilde{X}\to X$ is finite (and an isomorphism at all but a finite number of points). Therefore $\tilde{X}$ is defined by the sheaf $\pi_*\mathcal{O}_{\tilde{X}}=\prod\mathcal{O}_{X_i}$ of finite $\mathcal{O}_X$ algebras.

Since X is reduced, the map $\mathcal{O}_X \to \prod \mathcal{O}_{X_i}$ is injective. Hence also $\mathcal{O}_X^* \to (\prod \mathcal{O}_{X_i})^* = \pi_* \mathcal{O}_{\tilde{X}}^*$ is injective. The cokernels of these maps are concentrated at the points of intersection of the curves X_i. Now it is easily seen that $H^p(\tilde{X}, \mathcal{O}_{\tilde{X}}) \approx H^p(X, \prod \mathcal{O}_{X_i})$ and $H^p(\tilde{X}, \mathcal{O}_{\tilde{X}}^*) \approx H^p(X, \prod \mathcal{O}_{X_i}^*)$ for all p. Therefore we get surjective maps

$$H^1(X, \mathcal{O}_X) \to H^1(\tilde{X}, \mathcal{O}_{\tilde{X}}), \quad H^1(X, \mathcal{O}_X^*) \to H^1(\tilde{X}, \mathcal{O}_{\tilde{X}}^*)$$

and we want to relate the kernels of these maps. The methods used are well known (cf. [4], [5]) and so we sketch the results briefly. One can relate the curves X_i to their normalizations in a similar way.

We begin by constructing a curve X' as follows: Let $\mathcal{O}_{X'}$ the the subsheaf of $\prod \mathcal{O}_{X_i}$ whose sections on an open set U are those $f = (f_1, \cdots, f_s)$, $f_i \in \Gamma(U \cap X_i, \mathcal{O}_{X_i})$ such that $f_i(P) = f_j(P)$ for all $P \in U$ and all i, j with $P \in X_i \cap X_j$. $\mathcal{O}_{X'}$ is a sheaf of finite $\mathcal{O}_X$ algebras and hence a scheme X' and maps $\tilde{X} \to X' \to X$. Since $\mathcal{O}_{X',P}$ is actually a local ring for all $P \in X$, the underlying space of X' is homeomorphic to that of X. Clearly X' is maximal with these properties. The cokernel of the map $\mathcal{O}_{X'} \to \prod \mathcal{O}_{X_i}$ consists of vector spaces over k of dimension $n_P - 1$ concentrated at the points P of X where n_P is the number of components of X containing P, while the cokernel of $\mathcal{O}_{X'}^* \to \prod \mathcal{O}_{X_i}^*$ consists of abelian groups isomorphic to direct sums of $n_P - 1$ copies of k^* at P. With this in mind it is easily seen that the kernel of $H^1(\mathcal{O}_{X'}) \to H^1(\prod \mathcal{O}_{X_i})$ is a vector space of dimension $N = \sum_P (n_P - 1) - s + 1$ and that the kernel of $H^1(\mathcal{O}_{X'}^*) \to H^1(\prod \mathcal{O}_{X_i}^*)$ is isomorphic to k^{*N} (s is the number of components X_i).

Consider the map $\mathcal{O}_X \to \mathcal{O}_{X'}$. Its kernel is again concentrated at a finite set of points. Clearly $H^0(\mathcal{O}_X) \approx H^0(\mathcal{O}_{X'}) \approx k$. Hence the kernel of the map $H^1(\mathcal{O}_X) \to H^1(\mathcal{O}_{X'})$ is the direct sum of the groups $\mathcal{O}_{X',P}/\mathcal{O}_{X,P}$, $P \in X$. Similarly, the kernel of the map $H^1(\mathcal{O}_X^*) \to H^1(\mathcal{O}_{X'}^*)$ is the direct sum of the groups $\mathcal{O}_{X',P}^*/\mathcal{O}_{X,P}^*$ and we claim

LEMMA (1.6). *There exists an integer m, rings* $\mathcal{O}_{X',P} = \mathfrak{o}_0 \supseteq \mathfrak{o}_1 \supseteq \cdots, \supseteq \mathfrak{o}_m = \mathcal{O}_{X,P}$ and isomorphisms $\mathfrak{o}_{i-1}/\mathfrak{o}_i \approx \mathfrak{o}_{i-1}^*/\mathfrak{o}_i^*$, $i = 1, \cdots, m$.

Proof. Denote by $\mathfrak{o}'$, $\mathfrak{o}$ the rings $\mathcal{O}_{X',P}$, $\mathcal{O}_{X,P}$ and by $\mathfrak{m}'$, $\mathfrak{m}$ their maximal ideals respectively. Since $\mathfrak{o}'/\mathfrak{o}$ is a finite dimensional vector space over k one can show easily that the conductor of $\mathfrak{o}'$ in $\mathfrak{o}$ is $\mathfrak{m}$-primary (cf. [5], p. 67 ff.). Hence $\mathfrak{m}'^{m+1} \subseteq \mathfrak{m}$ for some integer m and so $\mathfrak{o} + \mathfrak{m}'^{m+1} = \mathfrak{o}$. Set $\mathfrak{o}_i = \mathfrak{o} + \mathfrak{m}'^{i+1}$, $i = 1, \cdots, m$. Then $\mathfrak{o}_i$ is a local ring and $\mathfrak{o}_0 = \mathfrak{o}'$, $\mathfrak{o}_m = \mathfrak{o}$. Moreover, denoting by $\mathfrak{m}_j$ the maximal ideal of $\mathfrak{o}_j$, we have $\mathfrak{m}_{i-1}^2 \subseteq \mathfrak{m}_i$. Clearly

$\mathfrak{m}_{i-1}/\mathfrak{m}_i \approx \mathfrak{o}_{i-1}/\mathfrak{o}_i$. Let $\mathfrak{m}_j^*$ be the subgroup of $\mathfrak{o}_j^*$ consisting of the elements of the form $1+x$, $x \in \mathfrak{m}_j$. Then also $\mathfrak{m}_{i-1}^*/\mathfrak{m}_i^* \approx \mathfrak{o}_{i-1}^*/\mathfrak{o}_i^*$.

Define $\phi\colon \mathfrak{m}_{i-1} \to \mathfrak{m}_{i-1}^*/\mathfrak{m}_i^*$ by $\phi(x) =$ residue class of $1+x$. Since

$$(1+x)(1+y)/(1+x+y) = 1 + (xy/1+x+y)$$

and $(xy/1+x+y) \in \mathfrak{m}_{i-1}^2 \subseteq \mathfrak{m}_i$ it follows that ϕ is a homomorphism. Clearly ϕ is surjective and its kernel is $\mathfrak{m}_i$. Thus $\mathfrak{m}_{i-1}/\mathfrak{m}_i \approx \mathfrak{m}_{i-1}^*/\mathfrak{m}_i^*$ as desired.

THEOREM (1.7). *For a positive cycle $Z = \sum r_i X_i$ the following are equivalent:*

(a) $H^1(\mathcal{O}_Z) = 0$.

(b) (1.2) *is an isomorphism, i.e.,* $H^1(\mathcal{O}_Z^*) \approx \mathbf{Z}^s$.

(c) *For every positive cycle $Y \leqq Z$, $p(Y) \leqq 0$.*

Proof. Since the equivalence of the first two statements is true for the curves X_i, it follows immediately for any cycle Z from the above discussion. We shall prove (a) $\Longleftrightarrow$ (c).

Let $0 < Y \leqq Z$ be given. Then the natural map $Y \to Z$ induces a map $H^1(\mathcal{O}_Z) \to H^1(\mathcal{O}_Y)$ which is surjective since X is 1-dimensional. Hence $\dim H^1(\mathcal{O}_Y) \leqq \dim H^1(\mathcal{O}_Z)$. Now certainly $\dim H^0(\mathcal{O}_Y) \geqq 1$, and so it follows from (1.1) that $p(Y) \leqq \dim H^1(\mathcal{O}_Y)$. Combining the two inequalities we get (a) $\Longrightarrow$ (c).

Assume (c) holds. In particular, $p(X_i) = 0$ for every component X_i of Z, i.e., X_i is a nonsingular rational curve. We use induction on $\sum r_i$, remarking that the theorem is true for $\sum r_i = 1$. Suppose therefore that the theorem is true for $\sum r_i - 1 > 0$. Fix a component X_0 of Z and set $Z_0 = Z - X_0$. Then (c) holds for Z_0 and so $H^1(\mathcal{O}_{Z_0}) = 0$ by the induction assumption. Hence if K is the kernel of the map $\mathcal{O}_Z \to \mathcal{O}_{Z_0} \to 0$ we have $H^1(K) \to H^1(\mathcal{O}_Z) \to 0$. Now K is isomorphic to the invertible sheaf $\mathrm{Tr}_{X_0}(-Z_0)$ (the cokernel of the map $\mathscr{L}(-Z) \to \mathscr{L}(-Z_0)$) as is seen from the diagram

(1.8)

$$\begin{tikzcd}
0 \arrow[r] & \mathscr{L}(-Z) \arrow[r] \arrow[d] & \mathcal{O}_V \arrow[r] \arrow[d] & \mathcal{O}_Z \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathscr{L}(-Z_0) \arrow[r] & \mathcal{O}_V \arrow[r] & \mathcal{O}_{Z_0} \arrow[r] & 0.
\end{tikzcd}$$

Since X_0 is a nonsingular rational curve it will follow that $H^1(K)=0$ and therefore $H^1(\mathcal{O}_Z)=0$ if $\deg(\mathrm{Tr}_{X_0}(-Z_0))=(-Z_0\cdot X_0)\geqq -1$. This last inequality is equivalent with $(Z\cdot X_0)\leqq 1+(X_0^2)$. Now, we may choose the component X_0 arbitrarily. Therefore we are done unless *for each* i $(Z\cdot X_i) > 1+(X_i^2)$. Suppose however that $(Z\cdot X_i)\geqq 2+(X_i^2)$ for each i. Since $p(X_i)=0$ we have $(K\cdot X_i)=-2-(X_i^2)$ by (1.1). Hence $(Z^2)+(Z\cdot K)=((Z+K)\cdot\sum r_iX_i)\geqq 0$ and so $p(Z)\geqq 1$. This contradicts (c) and completes the proof.

Section 2.

Assume $X=\cup X_i$ is connected and denote by $\|(X_i\cdot X_j)\|$ the intersection matrix. We recall the condition of Mumford [3]:

(2.1) If X is contractable then $\|(X_i\cdot X_j)\|$ is negative definite.

LEMMA (2.2). *Suppose X is contractable and let $\pi\colon V\to\bar{V}$ be its contraction. Then for any positive cycle Z with support on X, $\chi(\bar{V})-\chi(V)\geqq p(Z)$.* In particular, $\chi(\bar{V})\geqq\chi(V)$.

Proof. It follows from the Leray spectral sequence of the map π that $\chi(\bar{V})-\chi(V)$ is in this case equal to $\dim H^0R^1\pi_*\mathcal{O}_V$ where $R^1\pi_*\mathcal{O}_V$ is a sheaf concentrated at the image P of X on $\bar{V}$. The lemma follows easily from the theorem on holomorphic functions in Grothendieck's form [2]. For variety we argue as follows: $\chi(\bar{V})-\chi(V)$ is determined locally in a neighborhood of X on V and so we may assume V nonsingular. Then the surface $\bar{V}$ has only one singular point and hence is projective (cf. [6], Cor. II. 2. 6). Let $\bar{H}$ be a hypersurface section of $\bar{V}$ of sufficiently high degree so that $H^1(\mathscr{L}(\bar{H}))=H^2(\mathscr{L}(\bar{H}))=0$ and let H be the transform of $\bar{H}$ on V. We have an exact sequence $0\to\mathscr{L}(H-Z)\to\mathscr{L}(H)\to\mathcal{O}_Z\to 0$ and we may also assume $H^2(\mathscr{L}(H-Z))=0$. Then we find $\dim H^1(\mathscr{L}(H))\geqq\dim H^1(\mathcal{O}_Z)$. Therefore $\dim H^1(\mathscr{L}(H))\geqq p(Z)$ by (1.1). Now $\bar{H}$ is a locally principal divisor on $\bar{V}$, and so we may write the Riemann-Roch theorem [7] on $\bar{V}$ in the form $\dim H^0(\mathscr{L}(\bar{H}))=\chi(\bar{V})+\frac{1}{2}((\bar{H}^2)-(\bar{H}\cdot\bar{K}))$ (because $H^i(\mathscr{L}(\bar{H}))=0$, $i>0$). We also have, on V,

$$\dim H^0(\mathscr{L}(H))-\dim H^1(\mathscr{L}(H))=\chi(V)+\tfrac{1}{2}((H^2)-(H\cdot K)).$$

Clearly

$$\dim H^0(\mathscr{L}(\bar{H}))=\dim H^0(\mathscr{L}(H)) \text{ and } (\bar{H}^2)-(\bar{H}\cdot\bar{K})=(H^2)-(H\cdot K).$$

Therefore

$$\chi(\bar{V}) = \chi(V) + \dim H^1(\mathcal{L}(H)) \geqq \chi(V) + p(Z).$$

The last assertion is obtained by setting $Z = X_i$.

THEOREM (2.3). *With the above notations the following are equivalent:*

(a) *X is contractable and if $\pi\colon V \to \bar{V}$ is the contraction of X then* $\chi(\bar{V}) = \chi(V)$.

(b) i. *The intersection matrix $\| (X_i \cdot X_j) \|$ is negative definite.*

ii. *For every cycle $Z > 0$ with support on X, $p(Z) \leqq 0$.*[2]

Moreover, if V is a normal projective surface and (a) *holds, then $\bar{V}$ is also projective.*

Proof. (a) $\Longrightarrow$ (b) is contained in (2.1) and (2.2). Assume (b) holds. Since both assertions of (a) are local along X on V, we may assume V normal and projective, and to complete the proof we have to construct $\bar{V}$ projectively with $\chi(\bar{V}) = \chi(V)$.

Let G denote the free abelian group of divisorial cycles with support on X. For any cycle D on V define $f_D \in \operatorname{Hom}(G, \mathbf{Z})$ by $f_D(Y) = (Y \cdot D)$ for $Y \in G$. This yields in particular a homomorphism $i\colon G \to \operatorname{Hom}(G, \mathbf{Z})$. Since $\| (X_i \cdot X_j) \|$ is negative definite, i is an injection and since G and $\operatorname{Hom}(G, \mathbf{Z})$ have the same rank the cokernel of i is a finite group. Therefore for any cycle D on V we have, for suitable n and $Z \in G$, $nf_D = f_{nD} = -f_Z$.

Now let H be a hypersurface section of V of sufficiently high degree so that $H^1(\mathcal{L}(H)) = H^2(\mathcal{L}(H)) = 0$. Replacing H by a multiple if necessary, we may assume $f_H = -f_Z$ for some (unique) $Z \in G$. Since H is a hypersurface section, $(H \cdot X_i) > 0$ for all i, and hence since $\| (X_i \cdot X_j) \|$ is negative definite it follows that $Z > 0$, and support of Z is X. For, write $Z = A - B$ where $A, B \geqq 0$ are without common components. Since $(Z \cdot X_i) < 0$ for all i, $(A \cdot B) - (B^2) = (Z \cdot B) \leqq 0$. But $(A \cdot B) \geqq 0$ since A, B have no common components, and $(B^2) \leqq 0$ since $\| (X_i \cdot X_j) \|$ is negative definite. Hence $(Z \cdot B) \geqq 0$, i.e., $(Z \cdot B) = 0$, $B = 0$. Therefore $Z \geqq 0$. Clearly then every X_i must be a component of support of Z.

We have $f_{H+Z} = 0$ and so it is clear that the degree of $\operatorname{Tr}_Z(H + Z) = 0$.

---

[2] We remark that $p(Z) \leqq 0$ follows from $(b)i$ for 'sufficiently large' cycles Z. On the other hand, i is nearly a consequence of ii. The only case where ii holds but i does not is the case that X is set theoretically a 'ruling' on a ruled surface (not necessarily a relatively minimal model).

Therefore since (b) ii holds we can apply Theorem (1.7) to conclude that $\mathrm{Tr}_Z(H+Z) \approx \mathcal{O}_Z$. Thus we have an exact sequence

$$(2.4) \qquad 0 \to \mathcal{L}(H) \to \mathcal{L}(H+Z) \to \mathcal{O}_Z \to 0.$$

Now $H^1(\mathcal{L}(H)) = 0$ and hence the map $H^0(\mathcal{L}(H+Z)) \to H^0(\mathcal{O}_Z)$ deduced from (2.4) is surjective. Therefore there is a global section of $\mathcal{L}(H+Z)$ whose trace on Z is the constant 1, and so there is a divisor D in the linear system $|H+Z|$ which does not meet any point of $X =$ support of Z. Since H is a hypersurface section, any base point or fixed component of $|H+Z|$ must lie on X. Therefore $|H+Z|$ has no base points or fixed components and defines a map π' of V onto some projective surface V'. It is easily seen that π' is an isomorphism outside of X because $|H+Z| > |H|$ and H is ample. On the other hand, $((H+Z)\cdot X_i) = 0$ for all i. Since X is connected it follows that $\pi'(X)$ is a single point of V' and so if $\bar{V}$ is a normalization of V' then $\pi\colon V \to \bar{V}$ is the contraction of X. Since the normalization of a projective variety is again projective, $\bar{V}$ is also projective.

By Theorem (1.7), $H^1(\mathcal{O}_Z) = 0$. Since also $H^1(\mathcal{L}(H)) = 0$ it follows from (2.4) that $H^1(\mathcal{L}(H+Z)) = 0$. Denote by $\mathcal{L}(\bar{D})$ the invertible sheaf on $\bar{V}$ corresponding to $\mathcal{L}(H+Z)$. $\bar{D}$ is a locally principal divisor on $\bar{V}$ and $(\bar{D}^2) - (\bar{D}\cdot\bar{K}) = ((H+Z)^2) - ((H+Z)\cdot K)$. Since $H^2(\mathcal{L}(H)) = 0$, $H^2(\mathcal{L}(H+Z)) = 0$ and also $H^2(\mathcal{L}(\bar{D})) = 0$. On the other hand $H^1(\mathcal{L}(\bar{D})) \geqq 0$ and $\dim H^0(\mathcal{L}(H+Z)) = \dim H^0(\mathcal{L}(\bar{D}))$. Hence we get from the Riemann-Roch formula on V and $\bar{V}$, $\chi(\bar{V}) \leqq \chi(V)$ and so $\chi(\bar{V}) = \chi(V)$ by (2.2). This completes the proof.

COROLLARY (2.5). *In a given birational class of normal surfaces, a surface $\bar{V}$ with $\chi(\bar{V})$ minimal for the class is projective.*

A nonsingular surface is projective, so we may use induction on the number of singular points of $\bar{V}$. Choose for V a surface obtained by resolving the singularity at a particular point of $\bar{V}$. V may be assumed projective and we have $\chi(V) \geqq \chi(\bar{V})$ since $\chi(\bar{V})$ is minimal, hence $\chi(V) = \chi(\bar{V})$ by (2.2). Let X be the exceptional locus of the map $\pi\colon V \to \bar{V}$ and apply Theorem (2.3).

COROLLARY (2.6). *Suppose V normal. If X satisfies condition* (b) *of Theorem* (2.3) *and if D is any divisorial cycle on V with $(D\cdot X_i) = 0$ for all i then D is linearly equivalent to a cycle D_1 with $D_1 \cap X = \emptyset$.*

For, we may assume V nonsingular and projective. Then a hypersurface section H may be chosen so that $H^i(\mathcal{L}(H)) = H^i(\mathcal{L}(H+Z)) = 0$, $i > 0$, and so that, in the notations of the proof of (2.3), $f_H = -f_Z$, $Z \in G$. Then also $f_{H+D} = -f_Z$. Therefore we may conclude as above that there is a divisor A in $|H + Z|$ and a divisor B in $|H + D + Z|$ such that $A \cap X = B \cap X = \emptyset$. Take $D_1 = B - A$.

The following is a particularly simple case of (2.3) which seems however to be the most useful one. Let V be a normal surface, suppose X is contractable, and let $\pi\colon V \to \bar{V}$ be its contraction. If the canonical divisors on $\bar{V}$ are locally principal at the image P of X on $\bar{V}$ and if $\bar{K}$ is such a divisor then $\pi^{-1}(\bar{K})$ is defined. In general $\pi^{-1}(\bar{K})$ need not be a canonical divisor on V but we have

THEOREM (2.7). *With the above notations, the following are equivalent:*

(a) *X is contractable, $\bar{K}$ is locally principal at P and $\pi^{-1}(\bar{K})$ is a canonical divisor on V.*

(b) *$\|(X_i \cdot X_j)\|$ is negative definite, and every component X_i of X is a nonsingular rational curve with $(X_i^2) = -2$.*

Moreover, if (a) *or* (b) *holds, then* $\chi(V) = \chi(\bar{V})$.

The possible configurations of curves satisfying (b) are easily enumerated. They are depicted in Figure (2.8), where a line represents a curve and the intersections are transversal.

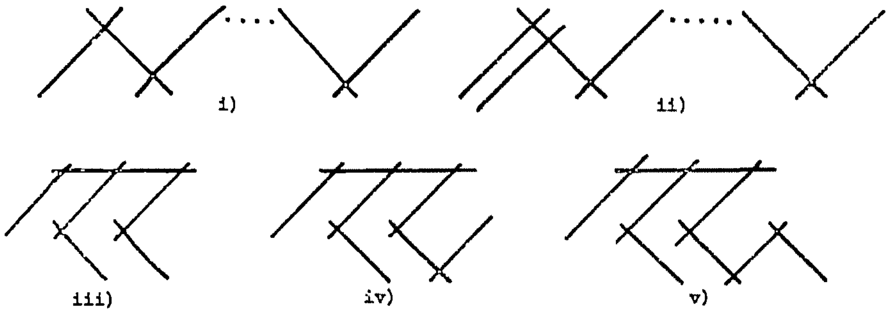

FIGURE (2.8).

Proof of (2.7). Suppose (a) holds. We may assume $\bar{K}$ does not pass through P. Then $\pi^{-1}(\bar{K})$ does not meet X. Hence $(X_i \cdot K) = 0$ for K canonical on V and for all i. Since $(X_i^2) \leqq -1$ we have $p(X_i) \leqq \frac{1}{2}$ by (1.2). Hence $p(X_i) = 0$ and $(X_i^2) = -2$. Since $\|(X_i \cdot X_j)\|$ is negative definite by (2.1) this shows (a) $\Longrightarrow$ (b).

Suppose (b) holds and let $Z=\sum r_iX_i$ be any positive cycle with support on X. Then $(Z^2)<0$. Since $(X_i^2)=-2$ and $p(X_i)=0$ we have $(X_i\cdot K)=0$ and so $(Z\cdot K)=0$. Hence $p(Z)<1$, i.e., $p(Z)\leqq 0$. Therefore condition (b) of (2.3) is satisfied and it follows that X is contractable and that $\chi(V)=\chi(\bar{V})$. Since $(X_i\cdot K)=0$ for all i it follows that there is a differential ω on V whose divisor does not meet X, by Corollary (2.6). Clearly then the divisor $(\omega)_{\bar{V}}$ of ω on $\bar{V}$ does not pass through P and $\pi^{-1}((\omega)_{\bar{V}})=(\omega)_V$. This completes the proof.

It seems unlikely that there is a stronger numerical criterion for contractability than that in (2.3) if the characteristic of k is zero, but if the characteristic is $p\neq 0$ one can do slightly better. Given Z, denote by D_Z the kernel of (1.2). For brevity we call a curve X *simply connected* if D_X contains a chain of subgroups $D_X=D_0\supseteq D_1\supseteq\cdots\supseteq D_n=0$ with $D_{i-1}/D_i\approx k$ for $i=1,\cdots,n$.

THEOREM (2.9). *Let X be a connected curve on V.*

(A) *Suppose the characteristic of k is not zero and that $\|(X_i\cdot X_j)\|$ is negative definite. Then X is contractable if X is simply connected. In particular, X is contractable if $p(X)=0$.*

(B) *Suppose k is the algebraic closure of a prime field of characteristic $p\neq 0$. Then X is contractible if and only if $\|(X_i\cdot X_j)\|$ is negative definite.*

Moreover, under the hypotheses of (A) *or* (B), *if $\pi\colon V\to\bar{V}$ is the contraction of X and V is a normal projective variety, so is $\bar{V}$.*

For the proof we will need the following

LEMMA (2.10). *Suppose $\|(X_i\cdot X_j)\|$ is negative definite. There is a positive cycle Y with support X such that for every $Z\geqq Y$, Z having also support X, the canonical may $H^1(\mathcal{O}_Z^*)\to H^1(\mathcal{O}_Y^*)$ is an isomorphism.*

Proof. Choose N sufficiently large so that $H^1(\mathcal{L})=0$ if $\deg(\mathcal{L})>N$ for any invertible sheaf $\mathcal{L}$ on X_i, all i, (cf. [4], Theorem 9). Since $\|(X_i\cdot X_j)\|$ is negative definite we can find a $Y>0$ with support X so that $(Y\cdot X_i)<-N+(X_i^2)$ for all i, and we claim this Y satisfies the conditions of the lemma.

Let Z be given with support X and $Z>Y$. Writing $Z-Y=\sum\alpha_iX_i$ $(\alpha_i\geqq 0)$ we have

$$((Z-Y)^2)=((Z-Y)\cdot\sum\alpha_iX_i)<0,$$

hence $((Z-Y)\cdot X_i)<0$ for some i with $\alpha_i>0$, say for $i=0$. Then $(Z\cdot X_0)<-N+(X_0^2)$. Set $Z_0=Z-X_0$. By induction on $\sum\alpha_i$ we may assume $H^1(\mathcal{O}_Z^*)\approx H^1(\mathcal{O}_Y^*)$, and we have to show that the map $H^1(\mathcal{O}_Z^*)\to H^1(\mathcal{O}_{Z_0}^*)$ is an isomorphism. But the kernels J, K of $\mathcal{O}_Z\to\mathcal{O}_{Z_0}$, $\mathcal{O}_Z^*\to\mathcal{O}_{Z_0}^*$ respectively are isomorphic as in (1.4) and hence $K\approx\mathrm{Tr}_{X_0}(-Z_0)$ by (1.8). Therefore the degree of K as a sheaf on X_0 is $(-Z_0\cdot X_0)=-(Z-X_0\cdot X_0)>N$, and so $H^1(K)=0$ by the assumption on N and the lemma follows from the exact cohomology sequence $H^1(K)\to H^1(\mathcal{O}_Z^*)\to H^1(\mathcal{O}_{Z_0}^*)\to 0$.

Proof of (2.9). To begin with, we note that condition (c) of (1.7) is verified for $Z=X$ if $p(X)=0$ since X is reduced and $p(X_i)=0$ for all i (necessarily $p(X)\geqq 0$ since X is also connected). This follows immediately from the formula $p(Z_1+Z_2)=p(Z_1)+p(Z_2)+(Z_1\cdot Z_2)-1$. Therefore (1.7)(b) implies that X is simply connected in this case and so the last assertion of (2.9)(A) follows from the first.

Let Y be a divisor as in Lemma (2.10). The reason the theorem is true is that D_Y (the kernel of (1.2)) is a torsion group under the assumptions of (A) or (B). In fact if X is simply connected then every element of D_Y has p-power order, $p=\operatorname{char}k$, since by the definition of simply connected and by Lemma (1.4) D_Y has a 'composition series' whose quotients are isomorphic to k-vector spaces. If k is the algebraic closure of the prime field of characteristic $p\neq 0$ then it is well known that each element of D_{X_i} has finite order. On the other hand, the kernel of the obvious map $D_Y\to\oplus D_{X_i}$ has by the discussion of § 1 a 'composition series' whose quotients are isomorphic to k-vector spaces or to k^*. Hence this kernel is a torsion group, and so is D_Y.

We try to copy the proof of (2.3): We may assume V normal and projective, and we choose a hypersurface section H so that $H^i(\mathcal{L}(nH))=0$ for $i,n>0$, and so that $f_H=-f_Z$ for some $Z\in G$ (notation as in the proof of (2.3)). The proof will go through if we can show that $\mathrm{Tr}_Z(H+Z)\approx\mathcal{O}_Z$ so as to obtain the exact sequence (2.4), and by Lemma (2.10) it suffices to show that $\mathrm{Tr}_Y(H+Z)\approx\mathcal{O}_Y$. Since $(H+Z\cdot X_i)=0$ for all i it follows that the cohomology class in $H^1(\mathcal{O}_Y^*)$ representing $\mathrm{Tr}_Y(H+Z)$ is at any rate in D_Y, and D_Y being a torsion group it follows that $\mathrm{Tr}_Y(n(H+Z))\approx\mathcal{O}_Y$ for some n. We may replace H by nH and Z by nZ. Then $\mathrm{Tr}_Y(H+Z)\approx\mathcal{O}_Y$ and so $\mathrm{Tr}_Z(H+Z)\approx\mathcal{O}_Z$ by (2.10) as required.

COROLLARY (2.11). *Every normal surface defined over the algebraic closure of the prime field of characteristic* $p\neq 0$ *is projective.*

The proof is the same as that of Corollary (2.5). Since a finite field is perfect, one can resolve the singularities of an algebraic surface over a finite field [1] and therefore obtain a nonsingular projective model dominating a given one. As the contraction of curves can be done starting with an arbitrary plane section H of high degree, it is not difficult to see that the projective embedding of a given normal surface may be obtained over any finite field of definition.

REFERENCES.

[1] S. Abhyankar, "On the field of definition of a nonsingular birational transform of an algebraic surface," *Annals of Mathematics*, vol. 65, No. 2 (1957).

[2] A. Grothendieck, *Éléments de Géométrie Algébrique*, Paris, 1960-.

[3] D. Mumford, "The topology of normal singularities of an algebraic surface and a criterion for simplicity," *Publications Mathematiques de l'Institut des Hautes Études Scientifiques*, No. 9 (1961).

[4] M. Rosenlicht, "Equivalence relations on algebraic curves," *Annals of Mathematics*, vol. 56, No. 1 (1952).

[5] J.-P. Serre, *Groupes algébriques et corps de classes*, Paris, 1959.

[6] O. Zariski, *Introduction to the problem of minimal models in the theory of algebraic surfaces*, Tokyo, 1958.

[7] ———, "Complete linear systems on normal varieties and a generalization of a lemma of Enriques-Severi, *Annals of Mathematics*, vol. 55, No. 3 (1952).

COMPLETE INTERSECTIONS AND CONNECTEDNESS.*

By ROBIN HARTSHORNE.[1]

0. Introduction. We will place ourselves in the context of Grothendieck's preschemes [2], since the results we obtain are valid over arbitrary (noetherian) ground rings. A closed subset V of a non-singular ambient prescheme X is a *complete intersection* if it is the intersection of $s =$ codimension of V in X hypersurfaces, where by a *hypersurface* we mean a closed subset of pure codimension 1 in X. Note that we do not care about the multiplicity of the intersection. In this paper we give a necessary topological condition for V to be a complete intersection in some ambient prescheme X. It is that V should be locally connected in codimension one, which means essentially that removing things of codimension bigger than one does not disconnect V.

In § 1 we make precise the concept of a topological space being connected in codimension k. In § 2 comes the heart of the proof, a property of local rings (Proposition 2.1) which is of some interests in itself. In § 3 we give the main theorem (Theorem 3.4) and various applications.

My thanks are due to R. C. Gunning, who asked whether the variety in Example 3.4.1 below was a complete intersection, and thus aroused my interest in the subject; also to A. Grothendieck for many helpful suggestions and conversations during the preparation of this paper.

1. Connectedness properties. Throughout this section, X will denote a notherian topological space (see [2], Ch. 0, § 2). We recall that if Y is an irreducible closed subspace of X, the *codimension of Y in X*, codim(Y,X), is the supremum of those integers r such that there exists a sequence of closed irreducible subspaces Z_i of X,

$$Y \leqq Z_0 < Z_1 < \cdots < Z_r \leqq X.$$

If Y is a closed subset of X, but no longer irreducible, we put

$$\operatorname{codim}(Y,X) = \inf\{\operatorname{codim}(Y',X)\}$$

---

* Received January 26, 1962.

[1] National Science Foundation Fellow.

where the infimum is taken over the closed irreducible subspaces Y' of Y. Note that by this definition, the codimension of the empty set is $+\infty$, being an infimum taken over an empty set.

PROPOSITION 1.1. *Let X be a noetherian topological space, and $k \geqq 0$ an integer. Then the following are equivalent:*

(i) *Whenever Y is a closed subset of X, and* $\operatorname{codim}(Y,X) > k$, *then $X - Y$ is connected.*

(ii) *If X' and X'' are irreducible components of X, then there is a* finite sequence $X' = X_1, X_2, \cdots, X_r = X''$ *of irreducible components of X, such that for each $i = 1, 2, \cdots, r-1$, $X_i \cap X_{i+1}$ is of codimension $\leqq k$ in X.*

Proof. (i) $\Rightarrow$ (ii). If (ii) is not satisfied, taking Y to be the union of all those subsets of X which are of the form $X_1 \cap X_2$, where X_1 and X_2 are irreducible components of X and $\operatorname{codim}(X_1 \cap X_2, X) > k$, we see that $X - Y$ is not connected, which contradicts (i).

(ii) $\Rightarrow$ (i). If (ii) is satisfied, and Y is a closed subset of X of codimension $> k$, then Y includes no set of the form $X_1 \cap X_2$ as above, and since the components of $X - Y$ are the sets $X_i \cap Y$, where X_i are the components of X, we see that $X - Y$ is connected.

Definition. If X satisfies the equivalent properties of Proposition 1.1, we say that it is *connected in codimension k.*

Remarks. 1.1.1. If X is connected in codimension k, then it is also connected in codimension k' for any $k' \geqq k$.

1.1.2. Being connected in codimension k, for any k, implies being connected.

1.1.3. Finally, if $k \geqq \dim X$ (recall that the *dimension of X* is the largest codimension of a non-empty closed subset), then X is connected in codimension k if and only if it is connected.

Example 1.1.4. A prescheme consisting of two planes which meet in a point is connected in codimension 2 but not connected in codimension 1.

We recall that if Y is a closed irreducible subset of the topological space X, a *generic point for Y* is a point $y \in Y$ such that Y is the closure of the set consisting of the single point y. We will say that X has the *generic point property* if every closed irreducible subset of X has a unique generic point. From now on we suppose that all spaces considered have the generic point property. (For example, any prescheme has this property.)

Definition. A *local topological space* is a topological space which has a single closed point. If X is any topological space and $y \in X$, we define the *local space of* X *at* y, X_y, to be the set of *generalizations of* y, that is, the set of points y' whose closure contains y.

LEMMA 1.2. *Let* X *be a connected topological space, and* Y *a closed subspace, such that for each* $y \in Y$, $X_y - y$ *is non-empty and connected. Then* $X - Y$ *is connected.*

Proof. We proceed by contradiction. Suppose $X - Y$ is the union of two disjoint non-empty relatively closed subsets X_1, X_2. Since for each $y \in Y$, $X_y - y$ is non-empty, Y contains no generic points of irreducible components of X, i.e. $X - Y$ is dense in X. Therefore, taking closures in X, we have $X = \bar{X}_1 \cup \bar{X}_2$, and $\bar{X}_1 \cap \bar{X}_2$ is non-empty, since X is connected. Let y be a generic point of a component of $\bar{X}_1 \cap \bar{X}_2$, and look at the local space X_y. I claim that $X_y - y$ is disconnected, which will be a contradiction, since y, being in $\bar{X}_1 \cap \bar{X}_2$, is necessarily in Y. Indeed, $X_y - y$ is the union of the relatively closed subsets

$$Z_1 = \bar{X}_1 \cap (X_y - y),$$

and

$$Z_2 = \bar{X}_2 \cap (X_y - y).$$

Now Z_1 and Z_2 are disjoint, since y was chosen to be a generic point of a component of $\bar{X}_1 \cap \bar{X}_2$. To see that they are non-empty, observe that since $y \in \bar{X}_1$, there is a generic point x_1 of a component of X_1 which is a generization of y. But certainly $x_1 \notin \bar{X}_2$, since Y_1 and X_2 are closed and disjoint in $X - Y$. Hence $x_1 \in Z_1$, and Z_1 is non-empty. Similarly one sees that Z_2 is non-empty.

PROPOSITION 1.3. *Let* X *be a noetherian topological space, and let* $k \geqq 0$ *be an integer. Then the following conditions are equivalent:*

(i) *For any* $y \in X$, X_y *is connected in codimension* k.

(ii) *Whenever* $y \in X$ *is such that* $\dim X_y > k$, *then* $X_y - y$ *is connected.*

Proof. (i) $\Rightarrow$ (ii) follows from the definitions.

(ii) $\Rightarrow$ (i). Let $y \in X$, and let Y be a closed subset of X_y, with codim $(Y, X_y) > k$. Then for any $z \in Y$, $\dim X_z > k$. (Note that localization is transitive: for any $z \in X_y$, $(X_y)_z = X_z$.) Therefore, by our hypothesis (ii), $X_z - z$ is connected, and certainly $X_z - z$ is non-empty, since $\dim X_z > k \geqq 0$. But X_y, being a local space, is connected, so we are in a position to apply the above lemma to the pair X_y, Y. We see that $X_y - Y$ is connected. Therefore X_y is connected in codimension k.

Definition. If X satisfies the equivalent properties of Proposition 1.3, we say that X is *locally connected in codimension* k (C_k for short).

Remarks. 1.3.1. C_k implies $C_{k'}$ whenever $k' \geqq k$.

1.3.2. C_k is a purely local property, hence we cannot expect to draw global conclusions from it. However, one sees easily that if X is connected and satisfies C_k, then it is connected in codimension k. The converse is not true (see the example below).

1.3.3. If $k \geqq \dim X$, C_k is the empty condition, since any local topological space is connected.

Example 1.3.4. From 1.3.2 above it follows that for a local topological space, C_k implies connected in codimension k. But they are not equivalent, since the property of being connected in codimension k is not stable under localization. For example, let X_0 be a variety in projective space $\boldsymbol{P}^4$ consisting of three planes A, B, C, such that A and B meet in a line; B and C meet in a line; but A and C meet only in a point x_0. Let $C(X_0)$ be the cone over X_0 in affine space $\boldsymbol{A}^5$, and let X be the local space of $C(X_0)$ at the origin. Then clearly X is connected in codim 1, since X_0 is, but X does not satisfy C_1. For if y is a generic point of the line $C(x_0)$ lying over x_0, then X_y looks like two planes meeting in a point, so is not connected in codimension 1.

2. Depth and Connectedness. We recall that if A is a noetherian local ring, a sequence of elements $x_1, \cdots, x_r$ of the maximal ideal $\mathfrak{m}$ is called an *A-sequence* if x_1 is not a zero-divisor in A, and for each $i = 2, \cdots, r$, the residue class of x_i modulo the ideal $(x_1, \cdots, x_{i-1})A$ is not a zero-divisor in the ring $A/(x_1, \cdots, x_{i-1})A$. One can show (see [1], [5], or [7] vol. II, Appendix 6) that this property is independent of the order of the x_i. One can also show that the number of elements in any A sequence is at most the dimension of A, and that all maximal A-sequences have the same number of elements. This enables us to make the following definition.

Definition. The *depth* (or *homological codimension*) of the local ring A is the number of elements in a maximal A-sequence. If the depth of A is equal to its dimension, we say that A is a *Cohen-Macaulay ring.*

Definition. If A is a noetherian ring and $k > 0$ is an integer, we say that A has the *property S_k of Serre* if for every prime ideal $\mathfrak{p}$ of A, the following is true:

1) if $\dim A_{\mathfrak{p}} \leqq k$, then $A_{\mathfrak{p}}$ is a Cohen-Macaulay ring;

2) if $\dim A_{\mathfrak{p}} > k$, then depth $A_{\mathfrak{p}} \geqq k$.

Remarks. 2.0.1. The property $S_{k'}$ implies S_k whenever $k' \geqq k$.

2.0.2. The property S_k is stable under localization by any multiplicative system of A.

2.0.3. For $k \geqq \dim A$, (where $\dim A = \sup \dim A_{\mathfrak{p}}$, the supremum being taken over all prime ideals $\mathfrak{p}$ of A) all the properties S_k are equivalent.

2.0.4. If A is a noetherian local ring of dim n, then S_n is equivalent to A being Cohen-Macaulay. Indeed, one implication is clear. The other follows from the fact ([7], vol. II, p. 399, Corollary 4), that if A is Cohen-Macaulay, so is $A_{\mathfrak{p}}$ for any prime ideal $\mathfrak{p}$ of A.

The following proposition, although elementary, is the keystone to the proof of our main results in this paper.

PROPOSITION 2.1. *Let A be a noetherian local ring, let $X = \operatorname{Spec} A$, and let $y = \{\mathfrak{m}\}$ be the closed point of X. If $X - y$ is disconnected, then the depth of A is at most* 1.

Proof. 1). By our hypothesis, we can write X as the union of two closed subsets Z_1, Z_2 whose intersection is y, and neither of which is equal to y. Let

$$(0) = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_s$$

be an irredundant primary decomposition of the zero ideal of A, where $\mathfrak{q}_i$ is primary for the prime ideal $\mathfrak{p}_i$. We may assume that none of the $\mathfrak{p}_i$ is the maximal ideal $\mathfrak{m}$, for if $\mathfrak{m}$ is an associated prime of (0), then depth $A = 0$ and we are done. With this assumption, the $\mathfrak{p}_i$ naturally fall into two subsets: those in Z_1 and those in Z_2. Neither subset is empty, for among the $\mathfrak{p}_i$ are all the minimal primes of A, i.e. the generic points of X. We now let

$$\mathfrak{a} = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_r, \qquad \text{where } \mathfrak{p}_1, \cdots, \mathfrak{p}_r \in Z_1,$$
$$\mathfrak{b} = \mathfrak{q}_{r+1} \cap \cdots \cap \mathfrak{q}_s, \qquad \text{where } \mathfrak{p}_{r+1}, \cdots, \mathfrak{p}_s \in Z_2.$$

Then $\mathfrak{a} + \mathfrak{b}$ is primary for $\mathfrak{m}$. Indeed, any prime ideal containing $\mathfrak{a} + \mathfrak{b}$ is in Z_1 and in Z_2, hence is $\mathfrak{m}$, so $\mathfrak{m}$ is the only associated prime of $\mathfrak{a} + \mathfrak{b}$, which means that $\mathfrak{a} + \mathfrak{b}$ is primary for $\mathfrak{m}$.

2). We now find elements $x, y \in A$ such that

(i) $x \in \mathfrak{a}$, $\qquad y \in \mathfrak{b}$,

(ii) $x \notin \mathfrak{p}_{r+1} \cup \cdots \cup \mathfrak{p}_s$, $y \notin \mathfrak{p}_1 \cup \cdots \cup \mathfrak{p}_r$,

(iii) $x \cdot \mathfrak{b} = (0)$, $y \cdot \mathfrak{a} = (0)$.

For each $j = r+1, \cdots, s$, $\mathfrak{a} \not\subseteq \mathfrak{p}_j$, since any prime ideal containing $\mathfrak{a}$ is in Z_1 and $\mathfrak{p}_j$ is not in Z_1. Therefore, by a well-known lemma on prime ideals, (see [7], vol. I, p. 215, Remark)

$$\mathfrak{a} \not\subseteq \mathfrak{p}_{r+1} \cup \cdots \cup \mathfrak{p}_s.$$

This shows that there exists an element x with the properties (i) and (ii) above. If $b \in \mathfrak{b}$, then $xb \in \mathfrak{a} \cdot \mathfrak{b} = 0$ and x has the property (iii). We choose y similarly.

3. We will now show that $x + y$ is a maximal A-sequence consisting of one element, and thus that depth $A = 1$.

a). From (i) and (ii) above it follows that each $\mathfrak{p}_i$ contains one and only one of the elements x, y, so that $x + y$ is in none of the $\mathfrak{p}_i$. Therefore $x + y$ is not a zero-divisor in A, and forms an A-sequence.

b). The element x is not in the principal ideal $(x + y)$. For suppose it were, say $x = a(x + y)$. Then for each $i = 1, \cdots, r$, $x \in \mathfrak{q}_i$, $x + y \notin \mathfrak{p}_i$, hence $a \in \mathfrak{q}_i$. Then $a \in \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_r = \mathfrak{a}$, so $ay = 0$ by (iii) above. But then $x = ax$, i. e. $x(1 - a) = 0$, which is impossible since $x \neq 0$ and $1 - a$ is a unit.

c). Next we observe that $x \cdot (\mathfrak{a} + \mathfrak{b}) \subseteq (x + y)$. Indeed, if $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$, then, noting that by (iii) we have $ya = xb = 0$, it follows that $x(a + b) = xa = (x + y)a \in (x + y)$.

d). Finally, we show that any element $z \in \mathfrak{m}$ is a zero-divisor in $A/(x + y)$, i. e. that $x + y$ is a maximal A-sequence. Since $\mathfrak{a} + \mathfrak{b}$ is primary for $\mathfrak{m}$, some power z^n of z is in $\mathfrak{a} + \mathfrak{b}$. Then by c) above, $xz^n \in (x + y)$. Let n be the least integer such that $xz^n \in (x + y)$. Then since $x \notin (x + y)$, $n \geqq 1$, and we can write $z \cdot xz^{n-1} \in (x + y)$, where $xz^{n-1} \notin (x + y)$, which shows that z is a zero-divisor in $A/(x + y)$. q. e. d.

THEOREM 2.2. *Let X be a connected, locally noetherian prescheme, and let Y be a closed subset of X such that for each $y \in Y$, the local ring $\mathcal{O}_{y,X}$ of y in X has depth at least 2. Then $X - Y$ is connected.*

Proof. By Lemma 1.2, we need only show that for each $y \in Y$, $X_y - y$ is non-empty and connected. From the definition of X_y, it follows that $X_y = \operatorname{Spec} \mathcal{O}_{y,X}$. Now since depth $\mathcal{O}_{y,X} \geqq 2$, we have $X_y - y$ connected by Proposition 2.1. Certainly it is non-empty, since $\dim X_y \geqq 2$.

COROLLARY 2.3. *Let X be a locally noetherian prescheme, and let $k \geqq 0$ be an integer. Suppose that for any $y \in X$ such that* $\dim \mathcal{O}_{y,X} > k$, *we have depth* $\mathcal{O}_{y,X} \geqq 2$. *Then X is locally connected in codimension k.*

Proof. Referring to criterion (ii) of Proposition 1.3, let $y \in X$ be such that $\dim X_y > k$. Then $\dim \mathcal{O}_{y,X} > k$, so by hypothesis we have depth $\mathcal{O}_{y,X} \geqq 2$. Applying the theorem to the pair X_y, y, we see that $X_y - y$ is connected. Therefore X is locally connected in codimension k.

COROLLARY 2.4. *If X is a locally noetherian prescheme, all of whose local rings have the property S_2 of Serre, then X is locally connected in codimension* 1.

Proof. The property S_2 of Serre implies that whenever $y \in X$ is such that $\dim \mathcal{O}_{y,X} > 1$, then depth $\mathcal{O}_{y,X} \geqq 2$. Hence we may apply the previous corollary with $k = 1$.

Remark 2.4.1. (Grothendieck) If A is a noetherian local ring having the property S_2, and if A satisfies the first chain condition of Nagata (namely that whenever $\mathfrak{p} \subseteq \mathfrak{p}'$ are prime ideals of A, then all maximal chains of prime ideals $\mathfrak{p} = \mathfrak{p}_0 < \mathfrak{p}_1 < \cdots < \mathfrak{p}_r = \mathfrak{p}'$ have the same length), then A is *equidimensional,* that is to say, all the irreducible components of $X = \operatorname{Spec} A$ have the same dimension. For by Corollary 2.4, X is connected in codimension 1; hence if X', X'' are irreducible components of X, there is a chain of irreducible components $X' = X_1, \cdots, X_r = X''$ such that for each $i = 1, \cdots, r-1$, $X_i \cap X_{i+1}$ has codimension at most 1. In fact, it must have codimension precisely 1 with respect to X_i and X_{i+1}, since it is equal to neither of them. Hence by the chain condition, $\dim X_i = 1 + \dim X_i \cap X_{i+1} = \dim X_{i+1}$. Iterating this argument r times we find $\dim X' = \dim X''$.

3. Application to complete intersections. Let X be a non-singular prescheme of dimension n (as for example projective or affine n-space over a field, or the spectrum of a regular local ring of dimension n), By *a hypersurface* in X we mean a closed subset H of X which is of pure codimension 1, i.e. such that the codimension of H at each point of H is 1.

Definition. A closed subset V of X, of codimension s, is a *complete intersection* in X if it can be written as the intersection of s hypersurfaces. It is *locally a complete intersection* if every point of V has a neighborhood in X in which V is a complete intersection.

Definition. A local ring A of dimension r is said to be a *complete inter-*

section if one can express it as the quotient of a regular local ring B by an ideal $\mathfrak{b}$, which can be generated by $\dim B - r$ elements.

It follows immediately from the Cohen-Macaulay theorem ([7], vol. II, p. 397, Theorem 2), that if $A = B/\mathfrak{b}$ is a complete intersection, as in the definition, then $\mathfrak{b}$ can be generated by a B-sequence, and A is a Cohen-Macaulay ring.

PROPOSITION 3.1. *Let V be a complete intersection in a non-singular prescheme X. Suppose given a particular representation*

$$V = H_1 \cap \cdots \cap H_s$$

of V as an intersection of $s = \operatorname{codim}(V, X)$ hypersurfaces, and suppose given also positive integers $n_1, \cdots, n_s$. For each $i = 1, \cdots, s$, let $\mathfrak{I}_i$ be the sheaf of ideals of H_i in $\mathfrak{O}_X$, that is, the sheaf of germs of sections of $\mathfrak{O}_X$ vanishing on H_i. Let $\mathfrak{J}$ be the sheaf of ideals

$$\mathfrak{J} = \mathfrak{I}_1^{n_1} + \cdots + \mathfrak{I}_s^{n_s},$$

and let $\mathfrak{O}_V$ be $\mathfrak{O}_X/\mathfrak{J}$. Then $(V, \mathfrak{O}_V)$ is a subprescheme of X, all of whose local rings are complete intersections. (We say that the prescheme structure $\mathfrak{O}_V$ is *induced* on V by the H_i taken with multiplicity n_i.)

Proof. Clearly $(V, \mathfrak{O}_V)$ is a subprescheme of X, since V is the variety of $\mathfrak{J}$. We need only check that the local rings are complete intersections. So let x be a point of V. Then the local ring $\mathfrak{O}_{x,V}$ of x on V is equal to the quotient $\mathfrak{O}_{x,X}/\mathfrak{J}_x$, where $\mathfrak{O}_{x,X}$ is a regular local ring. Hence to show that $\mathfrak{O}_{x,V}$ is a complete intersection it will be sufficient to show that the ideal $\mathfrak{J}_x$ can be generated by $t = \dim \mathfrak{O}_{x,X} - \dim \mathfrak{O}_{x,V}$ elements. Note that since $\mathfrak{O}_{x,X}$ is regular, t is also equal to $\operatorname{codim}_x(V, X)$.

Since H_i is of pure codimension 1 in X, and since $\mathfrak{O}_{x,X}$ is a unique factorization domain, the ideal $(\mathfrak{I}_i)_x$ of H_i at x is principal. Hence the ideal $\mathfrak{J}_x$ can be generated by s elements, and it will be sufficient to show that $s \leqq t$. But since the global codimension of V in X is s, the codimension t of V in X at a specific point x must be at least s. q. e. d.

Remark 3.1.1. One should note that the structure $\mathfrak{O}_V$ defined on V above is in general not the reduced induced prescheme structure on V. (See Example 3.4.5 below.)

Definition. We say that a closed subprescheme $(V, \mathfrak{O}_V)$ of a non-singular prescheme X is a *complete intersection* if

1) its underlying topological space V is a complete intersection in X, and

2) for some representation of V as a complete intersection of hypersurfaces, $V = H_1 \cap \cdots \cap H_s$, and for some choice of multiplicities n_i, $\mathcal{O}_V$ is the prescheme structure induced on V by this representation (in the sense of the above proposition).

PROPOSITION 3.2. *If the local ring A is a complete intersection, so is its completion A^*.*

Proof. Let A be written as a quotient $B/\mathfrak{b}$, where B is a regular local ring, and $\mathfrak{b}$ can be generated by $(\dim B) - (\dim A)$ elements. Taking completions, we have $A^* = B^*/\mathfrak{b}^*$, since completion is an exact functor ([7], vol. II, Ch. VIII, § 2). Now completion preserves the dimension and regularity of a local ring (ibid., Ch. VIII, § 11), so $\dim A = \dim A^*$; $\dim B = \dim B^*$, and B^* is regular. Moreover, $\mathfrak{b}^* = \mathfrak{b} \cdot B^*$, so $\mathfrak{b}^*$ can be generated by $(\dim B) - (\dim A) = (\dim B^*) - (\dim A^*)$ elements. Therefore A^* is a complete intersection.

LEMMA 3.3. *Let V be a closed subset of a noetherian prescheme X, and let $x \in V$. Let $\mathcal{O}_V$ be a prescheme structure on V making it into a subprescheme of X. Let V_x^* be the underlying topological space of* $\operatorname{Spec}(\mathcal{O}_{x,V})^*$, *where $\mathcal{O}_{x,V}$ is the local ring of x on V, and an asterisk denotes completion. Then V_x^* does not depend on the particular prescheme structure $\mathcal{O}_V$ chosen.*

Proof. Let $\mathcal{O}_V$ and $\mathcal{P}_V$ be two such structures. We may suppose that one of them, say $\mathcal{P}_V$, is the reduced induced structure on V. Then $\mathcal{P}_{x,V} = \mathcal{O}_{x,V}/\mathcal{N}$, where $\mathcal{N}$ is the ideal of nilpotent elements. Taking completions, we have

$$(\mathcal{P}_{x,V})^* = (\mathcal{O}_{x,V})^*/\mathcal{N}^*,$$

where $\mathcal{N}^* = \mathcal{N} \cdot (\mathcal{O}_{x,V})^*$ is an ideal consisting entirely of nilpotent elements. But dividing by nilpotent elements does not affect the set of prime ideals of a ring, so that underlying topological spaces of $\operatorname{Spec}(\mathcal{P}_{x,V})^*$ and $\operatorname{Spec}(\mathcal{O}_{x,V})^*$ are the same.

Definition. If V is a closed subset of a noetherian prescheme X, and $x \in V$, we call the topological space V_x^* defined above the *formal local space of V at x.* A local topological property of V will be said to hold *formally* if it holds for each formal local space of V.

THEOREM 3.4. *Let V be a closed subset of a non-singular prescheme X,* *which is locally a complete intersection. Then V is locally connected in codimension one, also formally.*

Proof. Since the question is local on V, we may suppose that it is globally a complete intersection in X. Then by Proposition 3.1 there is a prescheme structure $\mathcal{O}_V$ on V, all of whose local rings are complete intersections. But a local ring which is a complete intersection is also Cohen-Macaulay, and in particular, has the property S_2 of Serre (by Remarks 2.0.3, 2.0.4). Therefore, by Corollary 2.4, V is locally connected in codimension one.

As for the formal local spaces, V_x^* is the underlying topological space of $\operatorname{Spec}(\mathcal{O}_{x,V})^*$, and $(\mathcal{O}_{x,V})^*$ is a complete intersection by Proposition 3.2. Therefore the same argument shows that V_x^* is locally connected in codimension 1. q.e.d.

Remarks and Examples. 3.4.1. The simplest example to which our theorem applies is the subset of affine 4-space consisting of two planes which meet in a single point. For instance in $X = \boldsymbol{C}^4$, let V be the variety which is the union of the place where $z_1 = z_2 = 0$ with the place where $z_3 = z_4 = 0$. Then V, with the Zariski topology, is not connected in codimension 1 at the origin, hence is not a complete intersection. Note that by applying the theorem to the local ring of convergent power series at the origin, we see that V is not a complete intersection in the sense of several complex variables either.

3.4.2. Similarly, one can find an irreducible surface in $X = \boldsymbol{C}^4$ which is not a complete intersection. For example, let V be the surface given parametrically by

$$z_1 = t \qquad z_2 = tu$$

$$z_3 = u(u-1) \qquad z_4 = u^2(u-1).$$

Then one check easily that V is the one-to-one image of the (t,u)-plane, all except for the two points $(0,0)$ and $(0,1)$, both of which are mapped into the origin of X. Now V, being irreducible, is locally connected in codimension one. But formally at the origin, it looks like the first example above, so it is not formally connected in codimension one. Hence it is not a complete intersection.

3.4.3. The previous example sheds some light on the problem of how many hypersurfaces one must intersect to define an irreducible variety of dimension r in projective n-space. Martin Kneser has shown [3] that three hypersurfaces will suffice for curves in 3-space. One doesn't know if two will do for irreducible curves. Our example shows that two are not sufficient for irreducible surfaces in 4-space.

In this connection, Grothendieck asks if there is a complete normal local ring A of dimension 2 such that the underlying topological space Y of $\operatorname{Spec} A$ is not a complete intersection in any non-singular prescheme X. In dimension at least 3 there are such rings. For example, let $B = k[[x, y, z]]$ be the ring of formal power series over a field k of characteristic $\neq 2$, and let Γ be the automorphism sending x, y, z into $-x$, $-y$, $-z$ respectively. Then if $A = B^{\Gamma}$ is the ring of invariants of B under the action of Γ, the underlying topological space Y of $\operatorname{Spec} A$ is not a complete intersection in any X.

3.4.4. One might expect the following stronger formulation of our theorem to hold, namely that if V is a complete intersection, then any two irreducible components of V passing through a given point P have an intersection which is of codimension 1 with respect to each of them at P. This is not the case, however. For example, in $X = \boldsymbol{C}^4$, let V be the union of three planes Q_1, Q_2, Q_3, where Q_1 is defined by $z_1 = z_2 = 0$; Q_2 by $z_2 = z_3 = 0$; and Q_3 by $z_3 = z_4 = 0$. Then V is a complete intersection, defined by the two hypersurfaces $z_1z_3 + z_2z_4 = 0$ and $z_2z_3 = 0$. On the other hand, the two components Q_1 and Q_3 of V meet only in a point, which is of codimension 2.

3.4.5. A subprescheme of a non-singular prescheme X, whose underlying topological space is a complete intersection is not necessarily a complete intersection, even locally. For example, let V be the curve in affine 3-space X given parametrically by

$$x = t^3 \qquad y = t^4 \qquad z = t^5.$$

We will show that V, with the reduced induced structure $\mathcal{O}_V$, is not a complete intersection in the local space of X at the origin, although V is a set-theoretic complete intersection in X. Let $\mathcal{O}$ be the local ring of the origin in X, and let $\mathfrak{p} \subseteq \mathcal{O}$ be the ideal of V. Then to say that the prescheme $(V, \mathcal{O}_V)$ is a complete intersection in the local space X_0 of X at the origin, is to say that $\mathfrak{p}$ can be generated by two elements of $\mathcal{O}$. Now $\mathfrak{p}$ consists of all rational functions $f(x, y, z)/g(x, y, z)$ where f and g are polynomials over the base field k; $g(0, 0, 0) \neq 0$, and $f(t^3, t^4, t^5) \equiv 0$ as a polynomial in t. Clearly any such f can have no constant term or linear terms in x, y, z, so that $\mathfrak{p}$ must be contained in $\mathfrak{m}^2$. Reducing modulo $\mathfrak{m}^3$, we consider the k-vector space $\mathfrak{p}/\mathfrak{m}^3 \subseteq \mathfrak{m}^2/\mathfrak{m}^3$. Now $\mathfrak{p}$ contains the polynomials $y^2 - xz$, $x^3 - yz$, $z^2 - x^2y$, whose residues $\operatorname{mod} \mathfrak{m}^3$ are linearly independent over k. Hence the vector space $\mathfrak{p}/\mathfrak{m}^3$ has dimension at least 3. This shows that $\mathfrak{p}$ cannot be generated as an ideal by two elements, because if it could, their residues $\operatorname{mod}\mathfrak{m}^3$ would span $\mathfrak{p}/\mathfrak{m}^3$ as a k-vector space.

So we have seen that the prescheme $(V, \mathcal{O}_V)$ is not a complete intersection at the origin. On the other hand, the subset V is a complete intersection in X, given by the two hypersurfaces $z^2 - x^2y = 0$, and $x^4 + y^3 - 2xyz = 0$.

3.4.6. It is well known that a complete intersection of dimension at least 1 in projective space is connected. (This follows for example from the fact ([4], no. 77, 78), that a complete intersection V of dimension r in projective space is "r fois de première espèce." If r is at least one, we have $H^0(V, \mathcal{O}_V) = k$ which shows that V is connected.) Combining this with our theorem and the above Remark 1.3.2, we see that a complete intersection in projective space is (globally) connected in codimension one.

By a remark of Serre, one can also arrive at this result by a purely geometric argument in projective space: Suppose V is a complete intersection of dimension r in projective n-space, and suppose V not connected in codimension one. Then we can find closed subvarieties V' and V'' of V such that $V' \cup V'' = V$, and $\dim V' \cap V'' \leqq r-2$. Let W be a generic intersection of $n-r+1$ hyperplanes, and consider $W \cap V$. By the dimension theorem, we see that $W \cap V$ falls into two disjoint pieces, $W \cap V'$ and $W \cap V''$, each of dimension one. But $W \cap V$ is itself a complete intersection, and is of dimension 1, so must be connected, which is a contradiction.

HARVARD UNIVERSITY.

REFERENCES.

[1] Maurice Auslander and David Buchsbaum, "Codimension and multiplicity," *Annals of Mathematics*, vol. 68, no. 3 (1958), pp. 625-657.

[2] Alexandre Grothendieck, *Éléments de Géométrie Algébrique*, Institut des Hautes Etudes Scientifiques, Paris, 1960, 1961.

[3] Martin Kneser, "Ueber die Darstellung Algebraischer Raumkurven als Durchschnitte von Flächen," *Archiv für Mathematik*, vol. XI (1960), p. 157.

[4] Jean-Pierre Serre, "Faisceaux Algébriques Cohérents," *Annals of Mathematics*, vol. 61, no. 2 (1955), pp. 197-278.

[5] ———, *Algèbre Locale—Multiplicités*, mimeographed notes, Collège de France, 1957-58.

[6] ———, "Sur la dimension homologique des anneaux et des modules noethériens," *Proceedings of the International Symposium on Algebraic Number Theory*, Tokyo-Nikko, 1955.

[7] Oscar Zariski and Pierre Samuel, *Commutative Algebra*, in two volumes, Van Nostrand, 1958, 1960.

UNIQUE DIRECT INTEGRAL DECOMPOSITIONS ON CONVEX SETS.*[1]

By L. H. Loomis.

1. Introduction. In this paper we shall study a kind of commutative structure relative to an element of a general compact convex set K, and shall show that given any $z \in K$ and any maximal z-commutative structure B there exists a unique measure μ on K which is "compatible" with B and has z as its *resultant* or *expected vector*, $z = \int_K x d\mu(x)$. This unique B-measure μ turns out to be on the set E of extreme points of K to the extent that this is possible. This means that in the separable situation each such measure μ is supported by E, so that we obtain on the one hand another proof of the Choquet existence theorem [3], and that μ is always on E in the sense of the generalized existence theorem of Bishop and de Leeuw [1]. On the other hand, the uniqueness of μ relative to B can be considered to be a purely convex-theoretic generalization of the Mautner-Godement-Segal direct integral theorem of operator algebra theory [4, 6, 7]. In part, this theorem asserts the direct integral decomposability of a cyclic representation into irreducible parts, a result which appears as a special case of the later Choquet theorem when it is approached through the equivalent positive functionals. But it has the distinctive feature that the measure μ is obtained in a canonical way relative to a certain maximal commutative algebra of operators B.

Our notion of commutativity is of necessity very weak, and consequently we obtain *more* extreme measures representing a given element than arise from the earlier methods in either of the above situations. However, we shall show in Section 8 that if the partial ordering of measures used by Bishop and de Leeuw is replaced by a stronger partial ordering obtained by requiring simultaneous inequality over a whole class of orderings related to theirs, then a measure is maximal with respect to the simultaneous ordering if and only if it is maximal in our sense.

A formally stronger ordering (based on another notion of commutative system) is investigated briefly in Section 6. Although somewhat more cumber-

---

* Received March 7, 1962.

[1] Work on this manuscript was supported (in part) by the Army Research Office (Durham).

some to set up, it turns out to be optimal in the sense that its maximal elements (called s-maximal) are *exactly* the extremal measures. A strengthened version of Choquet's simplex theorem follows as a corollary. For the rest, the operator algebra situation is discussed in Section 7, and the paper concludes in Section 9 with a simple discussion of the situation occurring when K is a cone.

2. Preliminary considerations and definitions. It is assumed, of course, that the topology of K is reasonably related to its convex structure; specifically, we assume that there are on K enough continuous, real-valued, convex-linear functions to separate points. This is equivalent to assuming that K can be imbedded in the conjugate space of a real Banach space, for the set L of all such functions is a Banach space under the uniform norm and K is naturally identified with a subset of the closed unit sphere of the conjugate space L^*, its topology being that induced by the weak *-topology of L^*. Moreover, K then lies in the hyperplane $l_0(x)=1$, where l_0 is the constant function with value 1.

It is convenient now to enlarge K to the truncated cone C which K subtends at the origin, the new conical C then having K as its compact base. This is the setting of the problem used by Choquet, and, implicitly, by Bishop and de Leeuw, the issue now being the integral decomposition of points of the base K of the cone C. The chief advantage in enlarging K to C is that convex domination in K can be described in C in terms of the natural partial ordering of the cone. The convex combination $z=\alpha u+\beta v$ can now be written $z=x+y$ $(x=\alpha u, y=\beta v)$ and this relationship between z and x can be replaced by the inequality $x<z$, defined to mean that $z-x\in C$. We shall also want the related inequality between measures on K: $\nu<\mu$ if and only if $\mu-\nu$ is non-negative.

Let M be the set of non-negative regular Borel measures on K such that $\mu(K)\leqq 1$. Then M is also a conical convex set, and M is compact in the weak topology defined by the continuous functions on K. Given $\mu\in M$ there is a unique $x\in C$ determined by the condition that $l(x)=\int_K l(y)\,d\mu(y)$ for every $l\in L$, and we naturally express this relationship between μ and x by $x=\int_K y\,d\mu(y)$. We call x the *expected vector* or *resultant* of μ, and write $x=R(\mu)$. The set of all $\mu\in M$ such that $x=R(\mu)$ will be designated M_x. If $x\neq 0$ then M_x contains a point mass at the base of the segment through 0 and x, and M_x contains additional measures if and only if this base point is not an extreme point of K. In any case, M_x is a compact convex subset of M, and $\| \mu \| = \mu(K) = l_0(x)$ for all $\mu \in M_x$, l_0 being the linear functional with constant value 1 on K.

If A is a Borel subset of K and $0 < \delta \leqq 1$, let $K(A, \delta)$ be the set of all $x \in K$ such that there exists $\mu \in M_x$ satisfying $\mu(A) \geqq \delta$. Then $K(A, \delta)$ is clearly a convex subset of K including A, and is compact if A is compact. No point $x \in K(A, \delta) - A$ can be an extreme point of K since M_x contains more than one measure. We shall call any such difference set $K(A, \delta) - A$ with A compact an *elementary non-extreme set.*

A *subdivision* of an element $z \in C$ is a finite sequence of elements of C (repetitions being allowed) whose sum is z. The ordering of the sequence is of no consequence but the possibility of multiplicity is important. A subset $B \subset C$ will be called *z-commutative* if and only if it contains z and for every finite subset F of B there is a subdivision of z the set of whose partial sums includes F. If the subdivision can always be chosen with range in B then B will be called a *Boolean set on z.*

The family of z-commutative sets is partially ordered by inclusion, and the union of any linearly ordered subfamily is clearly itself z-commutative. It follows, by Zorn's lemma, that *every z-commutative set can be extended to a maximal z-commutative set.* The same holds for Boolean sets on z.

Any linearly ordered set of subelements of z which contains z is z-commutative, for if $0 = x_0 < x_1 < \cdots < x_n = z$, then the set $z_i = x_i - x_{i-1}$, $i = 1, \cdots, n$, is a subdivision having the required property. An element a will be said to commute with a z-commutative set B if and only if $B \cup \{a\}$ is z-commutative.

Given $\mu \in M$, the mapping $\nu \rightarrow R(\nu)$ is a continuous, convex-linear homomorphism of the set of measures $\nu < \mu$ onto a compact convex subset B_μ of subelements of $z = R(\mu)$. The set of these measures ν is the interval $[0, \mu]$ in the vector lattice of measures on K, and an easy induction shows that any interval $[0, a]$ in a vector lattice is a Boolean set on a. It follows that B_μ is a Boolean set on z. We shall say that a measure $\mu \in M$ is *compatible* with a z-commutative set B if $\mu \in M_z$ and $B \subset B_\mu$.

3. Existence.

THEOREM 1. *For every z-commutative set B there exists a measure $\mu \in M$ which is compatible with B.*

Proof. For each $x \in B$ let N_x be the set $M_x + M_{z-x}$. Then N_x is a compact convex subset of M_z. Moreover, the family $\{N_x \colon x \in B\}$ has the finite intersection property, for if F is any finite subset of B, if $\{z_1, \cdots, z_n\}$ is any subdivision of z the set of whose partial sums includes F, and if $\mu_i \in M_{x_i}$ for $i = 1, \cdots, n$, then $\mu = \sum_{1}^{n} \mu_i$ is in the intersection $\bigcap_{x \in F} N_x$. Therefore there exist $\mu \in \bigcap_{x \in B} N_x$ and clearly $B \subset B_\mu$.

COROLLARY 1. *Any z-commutative set can be extended to a Boolean set on z.*

COROLLARY 2. *If B is a maximal z-commutative set then $B = B_\mu$ for at least one $\mu \in M_z$.*

It is important to notice that the proof of the theorem goes through in the presence of further restrictions on the measures involved. In particular we have the following theorem.

THEOREM 1′. *Suppose that a Boolean set B on z is a subset of the cone based on $K(A, \delta)$ for some compact subset $A \subset K$ and some positive $\delta \leqq 1$. Then there exists a B-compatible measure μ satisfying $\mu(A) \geqq \delta\mu(K)$.*

The proof of this theorem is the same except that the sets M_x are replaced throughout by the sets M'_x consisting of those $\mu \in M_x$ such that $\mu(A) \geqq \delta\mu(K)$, and the subdivision of z is chosen with range in B.

COROLLARY. *If ν is a measure in M which is supported by $K(A, \delta)$ then there exists a B_ν-compatible measure $\mu \in M$ satisfying $\mu(A) \geqq \delta\mu(K)$.*

Proof. B_ν is a Boolean set on $R(\nu)$ and is a subset of the cone based on $K(A, \delta)$.

4. Uniqueness.

LEMMA 1. *Let P be the polynomial $(x_1 + x_2 + \cdots + x_n)^n$ and let Q be any polynomial obtained from P by changing the sign of one or more of the variables. Then the linear space spanned by P and all such polynomials Q contains the polynomial $x_1 x_2 \cdots x_n$.*

Proof. Let a be the coefficient of $x_1 x_2 \cdots x_n$ in P and let b be the coefficient of some other term T which contains at least one odd power and which necessarily omits some variable. That we have information about a and b is immaterial. Let P' be the polynomial obtained from P by adding the $n(n-1)/2$ polynomials Q in which exactly two signs are changed. P' is clearly symmetric. The important fact, however, is that in forming this sum the term T will occur sometimes positively and sometimes negatively, and that therefore its new coefficient is mb for some integer m less than

$n(n-1)/2$ in absolute value. The new coefficient of $x_1x_2\cdots x_n$ is of course $n(n-1)a/2$. It follows that there is a linear combination P_1 of P and P' which contains an $x_1x_2\cdots x_n$ term but no T term, nor any term which is a permutation variant of T. Treating P_1 exactly as we treated P above, for any other term T_1 containing at least one odd power, we can find a linear combination P_2 of P_1 and P_1' which retains an $x_1x_2\cdots x_n$ term but omits the symmetric polynomial generated by T_1. P_2 is of course a linear combination of P and certain of the Q polynomials. Continuing in this way we arrive at a polynomial P_j containing an $x_1x_2\cdots x_n$ term but having no other terms containing odd powers. Thus, if n is odd, P_j is just $cx_1x_2\cdots x_n$ and we are finished.

If n is even, then P_j consists of an $x_1x_2\cdots x_n$ terms plus terms containing only even powers. If $n=2$ this is already the case for P itself. If we now change the sign of just one variable and subtract we obtain the desired one term polynomial.

Remark. The linear space of the lemma does not include all homogeneous symmetric polynomials of degree n. For example, if $n=5$, no linear combination of P and the Q's can separate $x_1^4x_2$ from $x_1^2x_2^3$, the coefficients of these terms always being in the same ratio whatever signs are changed.

Now let μ be a probability measure on $[0,1]$, i.e., a Borel measure such that $\mu([0,1])=1$. Then, of course, μ determines and is determined by an increasing function α which is uniquely specified except for its ambiguous values at jump discontinuities by $\alpha(t-)=\mu([0,t))$, $\alpha(t+)=\mu([0,t])$, $\alpha(1)=1$. If β is the inverse function α^{-1}, again defined uniquely except for ambiguities at its jumps, it is well known and easy to see that

$$\int_0^1 f\,d\mu=\int_0^1 f(\beta(s))\,ds$$

for any Borel function f. In particular, if I is an interval of the form $[0,t)$ or $[0,t]$ and f is replaced by $f\chi_I$, we have

$$\int_I f\,d\mu=\int_0^{\mu(I)} f(\beta(s))\,ds.$$

Finally, β is uniquely determined by the continuous, convex function $g(a)=\int_0^a \beta(s)\,ds$, and it is clear from the above equation that $g(\mu(I))=\int_I s\,d\mu(s)$, and that g is linear over the a interval corresponding to a point t of positive measure. We thus have the following lemma.

LEMMA 2. *If μ is a probability measure on* $[0,1]$ *and if I is any interval of the form* $[0,t)$ *or* $[0,t]$, $t \leqq 1$, *then the set of pairs* $\langle \mu(I), \int_I s\,d\mu(s)\rangle$, *filled in to be linear between the two pairs corresponding to a point t of positive measure, is the graph of a continuous, convex function g, and the mapping $\mu \to g$ is a one-to-one correspondence between the set of all probability measures on* $[0,1]$ *and the set of all continuous, convex functions g on* $[0,1]$ *such that* $g(0)=0$, $g'(0+) \geqq 0$ *and* $g'(1-) \leqq 1$.

THEOREM 2. *If $B_\mu = B_\nu$ for two measures μ and ν in M then $\mu = \nu$.*

Proof. We may as well suppose that $\mu(K) = \nu(K) = 1$. Moreover, it will suffice to consider only those $l \in L$ such that $0 \leqq l \leqq 1$ on K. Given any such l, let I be a subset of K of the form $I_t = \{x : l(x) < t\}$ or $\bar{I}_t = \{x : l(x) \leqq t\}$. Such a set clearly minimizes the integral of l over all subsets of the same μ measure. More generally, if $\mu' < \mu$ is such that $\mu'(K) = \mu(I)$ then $\int_K l\,d\mu' \geqq \int_I l\,d\mu$, and if $\mu(I_t) < \mu'(K) < \mu(\bar{I}_t)$ then $\int_K l\,d\mu' \geqq \int l\,d\mu + cl(t)$, where $c = \mu'(K) - \mu(I_t)$. These inequalities say exactly that $\int_K l\,d\mu' \geqq g(\mu'(K))$, where g is the unique convex function (of Lemma 2) which is specified by $g(\mu(I)) = \int_I l\,d\mu$ for all I, and is linear between the two points corresponding to a value of t of positive measure.

Now if $B_\nu \subset B_\mu$ then for each I the restriction of ν to I has the same resultant as some $\mu' < \mu$. This implies both that $\nu(I) = \mu'(K)$ and that $\int_I l\,d\nu = \int_K l\,d\mu'$. If h is the convex function (of Lemma 2) specified by $h(\nu(I)) = \int_I l\,d\nu$ the above equations combine with the inequalities of the first paragraph to imply that $h(a) \geqq g(a)$ at every point a not interior to an interval of linearity of h. Since this inequality then holds at both endpoints of an interval of linearity for h, and since g is convex throughout this interval, it follows that $g \leqq h$ everywhere. But the hypotheses of the theorem are symmetric in g and h, and the same argument implies that $h \leqq g$. Thus $h \equiv g$.

It now follows from Lemma 2 that $\mu(I) = \nu(I)$ for each set I of the form $\{x \in K : l(x) < t\}$ or $\{x \in K : l(x) \leqq t\}$. That is, μ and ν have the same marginal distributions along l, or, equivalently, $\int f(l(x))\,d\mu(x) = \int f(l(x))\,d\nu(x)$ for every Borel function f. We have normalized l above to restrict the discussion to the standard interval $[0,1]$. The results clearly hold for any $l \in L$.

We now apply Lemma 1. For any finite set $\{l_1,\cdots,l_n\}\subset L$ the above argument shows that $\int f d\mu = \int f d\nu$ for $f=(\sum_1^n l_j)^n$, and therefore, by Lemma 1, $\int f d\mu = \int f d\nu$ for every f in the algebra generated by the linear functions. By the Stone-Weierstrass theorem this algebra is dense in the algebra $C(K)$ of all continuous functions on K. Thus $\int f d\mu = \int f d\nu$ for every continuous f, and $\mu=\nu$ by the Riesz representation theorem.

Remark. The last paragraph and Lemma 1 constitute a proof that a probability distribution on a compact subset of a vector space is known if all of its marginal distributions are known. In a finite number of dimensions this fact follows immediately from the fact that a finite measure is uniquely determined by its Fourier transform, and is well known to probabilists. The infinite dimensional case can be deduced from this remark, but the above proof may have independent interest.

COROLLARY 1. *If B is a maximal z-commutative set then there exists a unique B-compatible measure.*

We shall call any such measure a *maximal measure.*

COROLLARY 2. *If μ is maximal and $\nu<\mu$ then ν is maximal.*

Proof. If ν is not maximal then there exists $\nu'\neq\nu$ such that $R(\nu')=R(\nu)$ and $B_\nu\subset B_{\nu'}$. If $\mu'=(\mu-\nu)+\nu'$ then it is easy to see that $\mu'\neq\mu$, $R(\mu')=R(\mu)$ and $B_\mu\subset B_{\mu'}$. The inclusion must be proper by Theorem 2, and this contradicts the maximality of μ.

5. Maximal measures are on E.

In this section we shall apply a general type of argument used by the earlier authors [1, 2, 3]. Theorem 4 is essentially a new proof of the theorem of Bishop and de Leeuw on Baire sets. The fact that E is a G_δ set in the separable situation seems first to have been noticed by Godement [4].

LEMMA 3. *If F is an elementary non-extreme subset of K, then $\mu(F)=0$ for any maximal μ.*

Proof. If ν is the restriction of μ to F then ν is maximal by Corollary 2 to Theorem 2. But F is of the form $K(A,\delta)-A$, and if $\nu\neq0$ then ν is not maximal, by the corollary to Theorem 1′. Thus $\nu=0$.

If $\{l_n\}\subset L$ is a sequence that separates the points of K then every non-extreme point is interior to a segment in K along which some l_n varies. This suggests the following procedure. Given any $l \in L$, we consider a compact set $A = \{x : l(x) \leqq a \text{ or } l(x) \geqq b\}$ where a and b are rational numbers such that $a < b$, and a related disjoint convex set D which is the intersection of $K(A, 1)$ with the set where $c \leqq l(x) \leqq d$, c and d being rationals satisfying $a < c < d < b$. By Lemma 3, $\mu(D) = 0$ for any maximal μ, and according to the remark at the beginning of this paragraph $K - E$ is a countable union of such sets D. We thus have the following theorem.

THEOREM 3. *If the topology of K is countably generated (i.e., if L is separable) then the set E of extreme points is a G_δ set and every maximal measure is on E.*

The next theorem states as much as we can say in the general situation. We first prove a necessary elementary lemma.

LEMMA 4. *Let T be a continuous convex-linear mapping of K onto a compact convex set K^0. Let F be a compact subset of K and let z be a point of K which is separated from $K(F, \delta)$ by T. Then the image point $z^0 = T(z)$ lies outside of $K^0(F^0, \delta)$.*

Proof. Given any measure ν^0 on K^0, its restriction to F^0 is the image under T of a measure on F having the same norm, and its restriction to $K^0 - F^0$ is the image under T of a measure of equal norm on K. Therefore, if $z^0 \in K^0(F^0, \delta)$ then there exists $z' \in T^{-1}(z^0)$ such that $z' \in K(F, \delta)$ and hence z is not separated from $K(F, \delta)$ by T. The lemma follows.

THEOREM 4. *If F is a G_δ subset of K containing no extreme points then F is included in a countable union of elementary non-extreme sets. Consequently, $\mu(F) = 0$ for every maximal μ.*

Proof. We have to show that if A_n is a sequence of compact subsets of K whose union includes E then every $z \in K$ lies in some $K(A_n, \delta)$, where δ is a positive rational. Supposing the contrary, and setting $B_n = \bigcup_1^n A_j$, there exists $z \in K$ lying outside all the compact sets $K(B_n, \delta)$ for all positive δ. Fixing δ, there exists for each n a finite subset of L which separates z from $K(B_n, \delta)$. Let $\{l_j\}_1^\infty$ be the sequence of all these separating functionals, and let T be the continuous map into a separable space defined by this sequence.

Lemma 4 implies that $T(z) = z^0$ lies outside of $K^0(B_n^0, \delta)$ for every n. On the other hand, it is well known that $E^0 = T(E)$ contains all the extreme points of K^0, so that $z^0 \in K^0(\bigcup_1^\infty B_n^0, 1)$ by Theorem 3. Thus there exists an n such that $z^0 \in K^0(B_n^0, \delta)$, a contradiction.

Corollary. *If F is a Baire subset of K containing no extreme points then $\mu(F)=0$ for every maximal μ.*

Proof. By the regularity of μ we may assume F to be compact. But a compact Baire set is a G_δ set and the theorem applies.

What we have actually shown in these two theorems is that any measure μ which is zero on all elementary non-extreme sets is on E in the separable case, and in general is on E in the weaker sense of being zero on any G_δ or Baire subset of $K-E$. It is therefore natural to call any measure which is zero on all elementary non-extreme sets an *extremal measure.* Lemma 3 then says simpy that every maximal measure is extremal. The distinction between these two notions is worth making since it is conceivable that there exist extremal measures that are not maximal. However, we shall consider in the next section a situation in which this discrepancy cannot occur.

6. The strong ordering. Our results so far, when restricted to measures in M, have concerned the study of the partial ordering defined by $B_\mu \subset B_\nu$. We shall now see that the same theory is available for a certain somewhat more complicated, but stronger, ordering relation. Its set of maximal measures, which we shall call s-maximal, therefore contains all the measures that are maximal in the old sense. Indeed, this new set is exactly as large as one would hope, for we shall prove not only that every s-maximal measure is extremal but also, conversely, that every extremal measure is s-maximal. A strengthened form of the Choquet simplex theorem will follow as a corollary.

It should be mentioned at the outset that the author knows of no s-maximal measure that is not maximal. Moreover, it can be proved for some types of convex sets that every extremal (hence s-maximal) measure is maximal. It is quite conceivable that the notions of maximality and s-maximality actually coincide, and this whole question is entirely up in the air.

The reader will recall from Section 2 the notion of a subdivision of an element z. A subdivision $\{a_1,\cdots,a_m\}$ of z will be called a refinement of the subdivision $\{b_1,\cdots,b_n\}$ of z if the set of integers $\{1,\cdots,m\}$ can be partitioned into n disjoint subsets $I_1,\cdots,I_n$ so that $b_j=\sum\{a_i : i\in I_j\}$, $j=1,\cdots,n$. An s-Boolean set on z is a set of subdivisions of z which contains for each pair of its elements a common refinement. It is clear, as before, that any s-Boolean set on z can be extended to a maximal one. It is also important to notice that the set of all subdivisions of the interval of measures $[0,\mu]$ is an s-Boolean set on μ. This is seen most easily by replacing these measures by their Radon-Nikodym derivatives with respect to μ, and hence replacing $[0,\mu]$ (isomorphically) by the vector lattice interval of μ-measurable functions f such that $0 \leqq f \leqq 1$. Then if $\{f_1, \cdots, f_m\}$ and $\{g_1, \cdots, g_n\}$ are any two subdivisions of 1, the set $\{f_i g_j\}$ is clearly a common refinement. If $\mu \in M$ and we take resultants in the above situation we obtain an s-Boolean set on the element $z = R(\mu)$ which we shall designate S_μ. Finally, if S is any s-Boolean set on z, and if $\mu \in M_z$ and $S \subset S_\mu$, then we shall say that μ is compatible with S.

The whole theory of Sections 3-5 is now available. For example, to show that any s-Boolean set S on z is compatible with some $\mu \in M_z$, it is only necessary to modify Theorem 1 by defining, for each subdivision $\Delta \in S$, N_Δ to be the full sum set $N_\Delta = \sum_{x \in \Delta} M_x$, a non-empty compact convex subset of M_z. If Δ is a common refinement of Δ_1 and Δ_2 it is clear that $N_\Delta \subset N_{\Delta_1} \cap N_{\Delta_2}$, and the theorem finishes by way of the finite intersection property as before.

The analogue of Theorem 1′ follows similarly, the hypothesis now being that the *union* of all the subdivisions in S is a subset of the cone based on $K(A,\delta)$. Therefore, if ν is a measure supported by $K(A,\delta)$ then there exists an S_ν-compatible measure $\mu \in M$ satisfying $\mu(A) \geqq \delta\mu(K)$. The uniqueness of a measure μ such that $S_\mu = S$ follows directly from Theorem 2, for B_μ is simply the union of the subdivisions in S_μ, and the hypothesis $S_\mu = S_\nu$ thus gives at once $B_\mu = B_\nu$ and hence $\mu = \nu$. Thus if S is a maximal s-Boolean set on z there exists a unique measure $\mu \in M_z$ compatible with S, and if μ is any such s-maximal measure and $\nu < \mu$ then ν is also s-maximal, as in Corollary 2 to Theorem 2. The analogue of Lemma 3 follows, and therefore *every s-maximal measure is extremal.* Before showing that we now also have the converse theorem, we prove a necessary lemma. This lemma allows the quantitative application of a kind of localization argument due to Choquet.

LEMMA 5. *Suppose that the functionals l, $l' \in L$ and positive numbers a, ϵ are so related that $|l'(x) - a| \leqq \epsilon a/2$ whenever $l(x) \geqq 0\,(x \in K)$. Then if $\mu \in M$ is such that $\int l d\mu \geqq 0$, it follows that $\mu(\{x\colon l'(x) \leqq 0\}) < \epsilon\mu(K)$.*

Proof. We first show that if m is the maximum value of l on K then $l'(x) \leqq 0$ implies $l(x) < -m(1-\epsilon)/\epsilon$. Suppose on the contrary that y is such that $l'(y) \leqq 0$ and $l(y) \geqq -m(1-\epsilon)/\epsilon$, and let l assume its maximum value m on K at y_0. Then $l((1-\epsilon)y_0 + \epsilon y) \geqq 0$. Therefore

$$l'((1-\epsilon)y_0 + \epsilon y) \geqq a - \epsilon a/2 \text{ (and also } m \geqq 0).$$

But

$$\begin{aligned} l'((1-\epsilon)y_0 + \epsilon y) &= (1-\epsilon)l'(y_0) + \epsilon l'(y) \leqq (1-\epsilon)(a + \epsilon a/2) \\ &= a - \epsilon a/2 - \epsilon^2 a/2. \end{aligned}$$

The last two inequalities yield the contradiction $\epsilon^2 a/2 \leqq 0$.

Thus $\mu(\{x\colon l'(x) \leqq 0\}) \leqq \mu(A)$ where $A = \{x\colon l(x) < -m(1-\epsilon)\epsilon/\}$. Now suppose that $\int l d\mu \geqq 0$. If $l \equiv 0$ or $\mu(A) = 0$ the result is trivial, so we can suppose that m and $\mu(A)$ are positive. Then

$$0 \leqq \int l d\mu = \int_A + \int_{A'} < -m(1-\epsilon)\mu(A)/\epsilon + m(\mu(K) - \mu(A))$$
$$= m(\mu(K) - \mu(A))/\epsilon,$$

and the lemma follows.

THEOREM 5. *Every extremal measure is s-maximal.*

Proof. We shall show that if m is extremal and μ is an s-maximal measure compatible with S_m then $m < \mu$. That $m = \mu$ then follows from the fact that $R(m) = R(\mu)$.

Accordingly let G be a Borel set and let A be any compact set disjoint from G. Then $m(G \cap K(A,1)) = 0$, since m is extremal and $G \cap K(A,1)$ is a subset of the elementary non-extreme set $K(A,1) - A$. Therefore, given ϵ, there exists a compact subset $F \subset G$ such that F is disjoint from $K(A,1)$ and $m(G - F) < \epsilon$. A finite set $\{l'_1, \cdots, l'_n\} \subset L$ separates F from $K(A,1)$, and we can affinely adjust the l'_j so that each $l'_j \leqq 0$ on $K(A,1)$ and at each point of F at least one $l'_j \leqq 1$. We consider a single one of these functionals, say l'_1, and the subset $F_1 \subset F$ where $l'_1 \geqq 1$. The set F_1 can be divided into a finite number of strips $F_1{}^i$ on which $a_i \leqq l'_1 \leqq a_{i+1}$, the numbers a_i being chosen so that $m\{x\colon l'_1(x) = a_i\} = 0$ and $a_{i+1} - a_i < \epsilon$.

As above, if $A_1{}^i$ is the closed set $\{x \in K\colon l'_1 \leqq a_i \text{ or } l'_1 \geqq a_{i+1}\}$, then $m(K(A_1{}^i, 1) \cap F_1{}^i) = 0$ since m is an extremal measure and $m(F_1{}^i \cap A_1{}^i) = 0$. Also as above, we replace $F_1{}^i$ (and, similarly, the corresponding sets $F_j{}^i$) by a compact subset $H_1{}^i$ disjoint from $K(A_1{}^i, 1)$ and such that $\sum_{i,j} m(F_j{}^i - H_j{}^i) < \epsilon$. If we set $H = \bigcup_{i,j} H_j{}^i$ we thus have $m(G - H) < 2\epsilon$. Continuing the repeated argument, $H_1{}^i$ is separated from $K(A_1{}^i, 1)$ by a finite set of linear functionals which we can affinely adjust so that each is negative on $K(A_1{}^i, 1)$, and so that at each point of $H_1{}^i$ at least one is positive. If l_1 is any one of them then $l_1 \geqq 0$ implies $a_i < l'_1 < a_{i+1}$. Therefore, if ν is any measure on K such that $\int l_1 d\nu \geqq 0$, Lemma 5 allows us to conclude that $\nu(K(A,1)) < \epsilon\nu(K)$.

We have now arrived at the following conclusion. Given an extremal measure m, a Borel set G, a compact set A disjoint from G, and $\epsilon > 0$, there exists a compact subset H of G such that $m(G - H) < 2\epsilon$ and a finite decomposition of H, which we shall designate here simply $H = \bigcup_1^n H_i$, such that any measure μ_i with the same resultant as the restriction m_i of m to H_i must satisfy $\mu_i(A) < \epsilon\mu_i(K)$.

We now extend the set $\{m_i\}$ to a subdivision of m by throwing in the restriction of m to H' as m_{n+1}. The assumption that $S_m \subset S_\mu$ implies that there exists a corresponding subdivision $\{\mu_i\}_1^{n+1}$ such that $R(m_i) = R(\mu_i)$, $i = 1, \cdots, n+1$. Then

$$\mu(K - A) \geqq \sum_1^n \mu_i(K - A) \geqq (1-\epsilon) \sum_1^n \mu(K) = (1-\epsilon) \sum_1^n m_i(K)$$
$$= (1-\epsilon) \sum_1^n m(H_i) = (1-\epsilon) m(H) \geqq (1-\epsilon)(m(G) - 2\epsilon).$$

This inequality holds for every ϵ. Thus $\mu(U) \geqq m(G)$ for every open set U including G and $m < \mu$, q. e. d.

One of Choquet's two principal results concerns the kind of truncated cone C which in a finite dimensional space is a simplex [2]. In the separable case, his theorem asserts that there exists a unique extremal measure in M_z for every z in the base K of C if and only if C is a lattice in its natural partial ordering. If C is a lattice but the topology of K is not countably generated, then it can only be concluded that there can be at most one measure in M_z which is on E. The following theorem seems to complete the picture satisfactorily.

THEOREM 6. *For any element $z \in K$ the following conditions are equivalent:*

(a) *The set of all subelements of z is z-commutative.*
(b) *The set of all subelements of z is a lattice.*
(c) *The set of all subdivisions of z is an s-Boolean set on z.*
(d) *There exists a unique maximal measure in M_z.*
(e) *There exists a unique extremal measure in M_z.*

Proof. If the set B of *all* subelements of z is z-commutative then B is the unique maximal z-commutative set and hence, by Theorem 2, there exists a unique maximal measure μ in M_z. Conversely, if the set of all subelements of z is not z-commutative then there exists a finite subset that is not z-commutative and hence a *smallest* finite subset A that is not z-commutative. If A is divided into two non-empty subsets then each is z-commutative and has a maximal z-commutative extension. The two maximal extensions are distinct since their union includes A but neither alone does. Thus the properties (a) and (d) are equivalent.

In particular, the existence of a unique maximal measure $\mu \in M_z$ implies that the range of the resultant mapping $\nu \rightarrow R(\nu)$ on the domain $[0, \mu]$ is the set of all subelements of z. We can now deduce, furthermore, that the set of all subdivisions of z is an s-Boolean set on z. For if $\sum_1^n z_i = z$ and $z_i = R(\mu_i)$, we have, setting $\nu = \sum_1^n \mu_i$ and remembering that $\mu_i < \mu$, that $\nu/n < \mu$, ν/n is maximal, ν is maximal and $\nu = \mu$ (since $R(\nu) = z$ and μ is unique). Thus $\{z_1, \cdots, z_n\} \in S_\mu$, and the set of all subdivisions of z is exactly S_μ, so that (d) $\Longrightarrow$ (c).

The above conclusion also implies that the resultant mapping $\nu \rightarrow R(\nu)$ is one-to-one on $[0, \mu]$. For if $R(\mu_1) = R(\mu_2)$ then $\{R(\mu_1), z - R(\mu_2)\}$ is a subdivision of z and therefore $\{\mu_1, \mu - \mu_2\}$ is a subdivision of μ. That is, $\mu_1 = \mu_2$. The same type of argument also shows that $R(\mu_1) < R(\mu_2)$ if and only if $\mu_1 < \mu_2$. Therefore $\nu \rightarrow R(\nu)$ is an order isomorphism of the lattice $[0, \mu]$ onto $[0, z]$, proving that $[0, z]$ is a lattice also. Thus (d) implies (b).

That (b) implies (a) follows from our earlier remark (Section 2) that a vector lattice interval $[0, a]$ is a Boolean set on a. It is also clear (from Lemma 3) that (e) implies (d). Finally (c) implies that there exists a unique s-maximal measure $\mu \in M_x$, so that (c) implies (e) by Theorem 5. We thus have the implication chains (a) $\Longleftrightarrow$ (d), (d) $\Longrightarrow$ (b) $\Longrightarrow$ (a) and (d) $\Longrightarrow$ (c) $\Longrightarrow$ (e) $\Longrightarrow$ (d), and the theorem follows.

7. The representation theorem. The situation encountered in the direct integral decomposition of a group representation is somewhat different in that there exist subelements of z having special properties and the results center around measures that are compatible with maximal Boolean sets of these special subelements. We now briefly study this situation.

We shall call x a *part* of z if $x < z$ and if there exists no non-zero y such that $y < x$ and $y < z - x$. Thus a part of z is a subelement x which is "disjoint" from $z - x$. If A is the representation algebra generated by a positive functional z on an operator algebra, then every subfunctional $x < z$ corresponds to an operator in the commutating algebra A', and a part of z corresponds to a projection.

We make three probably redundant assumptions about the parts of z, the third of which, a direct statement of the spectral theorem, is a strong way of saying that the proper parts of z generate all of its subelements.

P1. If x is a part of y and y is a part of z then x is a part of z.

P2. If $x < z$ and x commutes with a part e of z then x is a *unique* sum $x = x_1 + x_2$ of subelements of e and $z - e$ respectively,

In this situation, if $x_1 < \lambda e$ ($> \lambda e$) we shall say that $x < \lambda z$ ($> \lambda z$) on e.

P3. If $x < z$ then there exists a unique increasing one-parameter family $\{e_\lambda : 0 < \lambda \leqq 1\}$ of parts of z with which x commutes, such that

(a) $x < \lambda z$ on e_λ and $x > \lambda z$ on $z - e_\lambda$

(b) $l_0(e_\alpha) \uparrow l_0(e_\beta)$ as $\alpha \uparrow \beta$, where l_0 is the functional which has constant value 1 on K.

We first notice that if $\mu \in M_z$ is compatible with a z-commutative set of parts B then for each $a \in B$ any $\nu < \mu$ such that $R(\nu) = a$ must be the restriction of μ to a set A, for otherwise a and $z - a$ are not disjoint. Moreover, for the same reason A is uniquely determined by a, so that the correspondence between the elements of B and the measures $\nu < \mu$ of which they are the results is one-to-one. If the restrictions of μ to A_1 and A_2 have resultants in B then the resultant of the restriction of μ to $A_1 \cap A_2$ is a part of z by P1. Thus any maximal commutative set of parts of z is a Boolean algebra.

LEMMA 6. *If $\mu \in M_z$ is compatible with the spectral family $\{e_\lambda\}$ of an element $x < z$ then $x \in B_\mu$.*

Any increasing family $\{e_\lambda\colon 0 < \lambda \leqq 1\}$ of parts of z satisfying P3 (b) *is the spectral family of a uniquely determined $x < z$.*

Proof. As we have seen, if $\mu \in M_z$ is compatible with $\{e_\lambda\}$ then e_λ is the resultant of the restriction of μ to a uniquely determined set A_λ. Clearly $A_\alpha \subset A_\beta$ if $\alpha < \beta$; also $l_0(e_\alpha) = \mu(A_\alpha)$ so that $A_\beta = \bigcup_{\alpha<\beta} A_\alpha$ by P3 (b). Let f be the function uniquely determined by $f < \lambda$ on A_λ, $f \geqq \lambda$ on A'_λ, and $f = 1$ on A'_1. It is clear that if $d\nu = f d\mu$ and $x = R(\nu)$ then x has $\{e_\lambda\}$ as its spectral family. If $\{e_\lambda\}$ is also the spectral family of x' then, for any $\delta > 0$, there are subelements $r < s < z$ constructed from $\{e_\lambda\}$ as upper and lower Riemann sums such that $s - r < \delta z$, $r < x < s$ and $r < x' < s$. Since $\| x - r \| \leqq l_0(s - r) \leqq \delta l_0(z) \leqq \delta$, and similarly for x', we have $\| x - x' \| \leqq 2\delta$, and therefore, δ being arbitrary, that $x = x'$.

LEMMA 7. *A subelement x of z commutes with a Boolean set B of parts of z if and only if its spectral family commutes with B.*

Proof. If x commutes with B and $\{b_1, \cdots, b_n\}$ is a B-subdivision of z then x is uniquely a sum $x = \sum_1^n x_i$ with $x_i < b_i$. If $\{e^i{}_\lambda\}$ is the spectral family of x_i and $e_\lambda = \sum_i e^i{}_\lambda$ then $\{e_\lambda\}$ is clearly the spectral family of x. Thus the spectral family of x commutes with B. Conversely if the spectral family $\{e_\lambda\}$ of x commutes with B and $e^i{}_\lambda$ is the part of e_λ in b_i then $\{e^i{}_\lambda\}$ is increasing and must satisfy P3 (b) for each i in order for it to hold for $\{e_\lambda\}$ itself. If x_i is the element with spectral family $\{e^i{}_\lambda\}$ then $x_i < b_i$ and $x = \sum x_i$ by Lemma 6, so that x commutes with B.

THEOREM 7. *If B_0 is a maximal commutative set of parts of z then there exists a unique $\mu \in M_z$ compatible with B_0.*

Proof. Let μ be *any* measure in M_z compatible with B_0 and let B be *any* maximal z-commutative set including B_0. The spectral elements e_λ of any $x \in B$ commute with B_0 by Lemma 7 and therefore belong to B_0 because B_0 is maximal. Then $x \in B_\mu$ by Lemma 6. Thus $B \subset B_\mu$ and hence $B = B_\mu$. Thus both μ and B are uniquely determined.

That the canonical μ of the Mautner-Godement-Segal theorem is unique in the above sense was shown by Tomita [8].

8. Comparison theorems. Bishop and de Leeuw partially order the measures in M_x by the requirement that $\int_K l^2 d\mu \leqq \int_K l^2 d\nu$ for all $l \in L$, and show that any measure which is maximal in this partial ordering is extremal. We shall show below that any measure which is maximal in their sense is maximal in our sense, but that the converse does not hold. However, if l^2 is replaced by $\Phi(l)$ and the above inequality is required to hold for all $l \in L$ and all continuous convex Φ, then a measure is maximal in this stronger partial ordering if and only if it is maximal in our sense. The section concludes with an example of a maximal measure in the operator algebra context which is not of the classical type.

THEOREM 8. *If μ and ν are measures on M such that $B_\nu \subset B_\mu$ then $\int \Phi(l)\,d\nu \leqq \int \Phi(l)\,d\mu$ for every $l \in L$ and every continuous convex function Φ. If the inclusion $B_\nu \subset B_\mu$ is proper then there exists $l \in L$ such that $\int l^2 d\nu < \int l^2 d\mu$.*

Proof. Midway through the proof of Theorem 2 we saw that if $B_\nu \subset B_\mu$ and $l \in L$ then the corresponding increasing functions β_1 and β_2, determined as in the proof of Lemma 2 by the one-dimensional distribution of ν and μ along l, satisfy $h(a) = \int_0^a \beta_1 \geqq \int_0^a \beta_1 = g(a)$. Moreover if the equality always holds, for all $l \in L$, then $\mu = \nu$. Thus if $\mu \neq \nu$ and $B_\nu \subset B_\mu$ the strict inequality must hold for some l and some a.

By the basic identity in the proof of Lemma 2, $\int \Phi(l)\,d\nu \leqq \int \Phi(l)\,d\mu$ if and only if $\int_0^1 \Phi(\beta_1(t))\,dt \leqq \int_0^1 \Phi(\beta_2(t))\,dt$ and the theorem is thus reduced to the following lemma.

LEMMA 8. *If f and g are increasing non-negative function on $[0,1]$ such that $\int_0^1 f = \int_0^1 g$, then $\int_0^a f(t)\,dt \geqq \int_0^a g(t)\,dt$ for all a if and only if $\int_0^1 \Phi(f(t))\,dt \leqq \int_0^1 \Phi(g(t))\,dt$ for all continuous convex functions Φ. Moreover, if the first inequality is strict for some a then $\int_0^1 f^2 < \int_0^1 g^2$.*

The first part of the lemma is a theorem of Hardy, Littlewood and Pólya [5, p. 170]. The proof of the second statement follows easily and we omit it.

COROLLARY 1. *A measure which is maximal in the sense of Bishop and de Leeuw is maximal in the sense of this paper.*

Proof. The theorem shows that if ν is not maximal in our sense then it is not maximal in the sense of Bishop and de Leeuw.

COROLLARY 2. *The measure μ is maximal if and only if it is maximal in the partial ordering defined by the requirement that $\int \Phi(l)\,d\nu \leqq \int \Phi(l)\,d\mu$ for all $l \in L$ and all continuous convex functions Φ.*

Example 1. Let K be the convex plane figure obtained from a square by bulging out its sides to equal circular arcs with radius greater than half the square diagonal. Then K lies properly within the circle S circumscribed about the square and has with its circumference only the four square vertices in common. Let μ be the measure having mass 1/4 at each of the midpoints of the arcs of K. It is easy to see that μ is maximal in our sense. However μ is not maximal in the sense of Bishop and de Leeuw, being strictly dominated in their sense by the measure ν having mass 1/4 at each of the vertices of the square. The details will be omitted.

Example 2. If A is the representation algebra of an operator algebra (or any C^*-algebra) defined by a normalized positive functional z then there exist direct integral decompositions of A into irreducible representation algebras that are *not* obtained from maximal projection algebras in A'. To see this let A_1 and A_2 be any positive operators in A' that do not commute. We can suppose that $A_1 + A_2 \leqq I$ (multiplying by scalars if necessary). Then A_1 and A_2 corresponds to subfunctionals X_1 and X_2 of z such that $X_1 + X_2 < z$. The subdivision $\{X_1, X_2, z - (X_1 + X_2)\}$ of z can be extended to a maximal z-commutative set, and if μ is the unique compatible measure on the irreducible functionals then the resultants of the restrictions of μ to subsets cannot form a Boolean algebra of projections. The details, again, will be omitted.

9. The structure on K. It may happen that K is already conical in a Banach space and therefore *itself* carries a partial ordering. If K has a compact convex base K_1 then the above theory can be applied *within* K and yields the direct integral expression of a point z of the base K, with respect to a maximal z-commutative subset of K. For the points of K_1 we therefore apparently have two theories, one by way of C and one within K. However, we shall see shortly that these two theories are in fact one.

Of course, K can be conical without having a compact convex base. For example, the set K of those continuous positive definite functions ϕ on a locally compact group such that $\phi(e) \leqq 1$ is compact, convex and conical, and even has a convex base K_1 consisting of those $\phi \in K$ such that $\phi(e) = 1$, but K_1 is not compact in general.

Moreover, *any* compact convex set K can be made conical with respect to any one of its extreme points x_0 by replacing the space L of all real-valued, continuous, convex-linear functions on K by its subspace L_0 of deficiency one consisting of those $l \in L$ such that $l(x_0) = 0$. Then K imbeds conically in L_0^* since x_0 clearly becomes the 0 of L_0^*.

Since L_0 by definition is the subspace of L annihilated by x_0 as a functional of L^*, the conjugate space L_0^* is the quotient space obtained from L by identifying the points on lines parallel to the line through 0 and x_0. Thus K as a cone with origin point at x_0 is isomorphic to the fibering $[C]$ of C into its segments parallel to the line through 0 and x_0. When K has a compact convex base K' (with respect to x_0) then the conical set subtended by K' at 0, which might be called the face of C opposite the vertex x_0, is clearly a cone in L^* which is isomorphic to $K = [C]$ in L_0^*. Thus the decomposition of elements $z \in K'$ relative to the intrinsic partial order in the conical K subtended by K' at x_0 becomes the decomposition we have already treated in the cone subtended by K' at 0. The identification of these two conical sets is simply the identification of the opposite ends of the segments in the fibering $[C]$, the points of K' being self-identified.

The above isomorphism holds for the convex decompositions of a "base point" z even when K does not have a base. A base point is now defined as one that is self-identified in the above correspondence and is thus one that is the last point in K on a ray from x_0. The following lemma states the fundamental fact.

LEMMA 9. *If z is a base point of K with respect to x_0, i.e., if $z = [z]$, then $z = \alpha x + \beta y$ in C $(\alpha + \beta = 1,\ \alpha > 0,\ \beta > 0)$ if and only if $[z] = \alpha[x] + \beta[y]$ in $[C]$. In this case x and y are also base points of K.*

Proof. The proof is almost obvious. If $[z]=\alpha[x]+\beta[y]$ in $[C]$ then z and $\alpha x+\beta y$ are points of C on the same fiber (since $\alpha[x]+\beta[y]=[\alpha x+\beta y]$) and therefore $z=\alpha x+\beta y$, $[z]$ containing only z. Since this holds for any x in $[x]$ we clearly have $x=[x]$, and similarly for y.

Thus for a base point z of K our decomposition theory in C is entirely equivalent to the theory that would be obtained from the intrinsic ordering in K restricting that ordering to those subelements of z in K that occur as summands in convex decompositions of z. These will be all of the subelements of z in K if K has a convex, but still not necessarily compact, base, as in the example of the positive definite functions.

HARVARD UNIVERSITY.

REFERENCES.

[1] E. Bishop and K. de Leeuw, "The representation of linear functionals by measures on sets of extreme points," *Université de Grenoble, Annales de l'Institut Fourier*, vol. 9 (1959), pp. 305-331.

[2] G. Choquet, "Unicité des représentations intégrales au moyen de points extremaus dans les cônes convex reticulés," *Comptes Rendus Hebdomadaires des Séances d l'Académie des Sciences*, Paris, vol. 243 (1956), pp. 555-557.

[3] ———, "Existence des représentations intégrales au moyen de points extremaux dans les cônes convex," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, Paris, vol. 243 (1956), pp. 699-702.

[4] R. Godement, "Sur la théorie des représentations unitaires," *Annals of Mathematics*, (2) vol. 53 (1951), pp. 68-124.

[5] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press, 1934.

[6] F. I. Mautner, "Unitary representations of locally compact groups, I," *Annals of Mathematics*, (2) vol. 51 (1950), pp. 1-25.

[7] I. Segal, "Decompositions of operator algebras I, II," *Memoirs of the American Mathematical Society*, vol. 9 (1951).

[8] M. Tomita, "Harmonic analysis on locally compact groups," *Mathematical Journal of Okayama University*, vol. 5 (1956), pp. 133-193.

LIBRARY
UNIVERSITY

DISCRETE GROUPS, SYMMETRIC SPACES, AND GLOBAL HOLONOMY.*

By Joseph A. Wolf.[1]

1. **Introduction.** Let M be a connected simply connected Riemannian manifold and let Γ be a properly discontinuous group of isometries such that M/Γ is compact. If every sectional curvature of M is negative, in particular if M is a noncompact irreducible symmetric space of rank 1, then a method of É. Cartan shows that every abelian subgroup of Γ is either finite or the product of a finite group with an infinite cyclic group. If M is the Euclidean space R^n, then a calculation shows that every abelian subgroup of Γ is the product of a finite group with a free abelian group on $\leqq n$ generators. These phenomena are unified by one of the conclusions of our Theorem 6.2: *If M is Riemannian symmetric and v is the maximum of the dimensions of those totally geodesic submanifolds of M which are isometric to Euclidean spaces, then every abelian subgroup of Γ is the product of a finite group with a free abelian group on $\leqq v$ generators, and Γ has a subgroup which is free abelian on v generators.* We also prove that an abelian subgroup of Γ must preserve a flat connected totally geodesic submanifold of M; if M/Γ is a manifold, it follows that M/Γ contains a maximal connected flat totally geodesic submanifold which is closed, and every abelian subgroup of $\pi_1(M/\Gamma)$ can be represented by closed geodesic arcs lying in a connected flat totally geodesic submanifold (Corollary 6.6). In addition, we analyze the group of components of the homogeneous holonomy group of a locally symmetric Riemannian manifold N (Theorem 7.1), prove that N has compact homogeneous holonomy group if N is compact[2] (Corollary 7.2), and give conditions for every manifold locally isometric to N to have compact homogeneous holonomy group (Corollary 7.3).

Our bounds are obtained by estimating the "size" of abelian subgroups of discrete-uniform subgroups of Lie groups $L = E \times G$ where E is a semi-direct product of a compact group and a vector group, such as the Euclidean group, and G is a reductive Lie group with only finitely many components.

---

* Received October 26, 1961; revised August 4, 1962.

[1] The author thanks the National Science Foundation for fellowship support during the preparation of this paper.

[2] If N is flat, this is just the classical Bieberbach Theorem [3]. This does not give a new proof of the Bieberbach Theorem because that result is used in our arguments.

The estimates are first made for reductive groups (Theorem 4.2), and then extended by a generalization of Bieberbach's Theorem (Theorem 5.1).

Let Γ be a discrete uniform subgroup of a reductive Lie group G, where G has only finitely many components, and let Δ be an abelian subgroup of Γ. Our main idea is that the size of Δ can be estimated by finding a Cartan subgroup H of G which is normalized by Δ, and observing that $\Delta \cap H$ has finite index in Δ. In order to find H, we first prove that every element of Γ is a semisimple element of G (Theorem 3.2), and then apply a result of A. Borel and G. D. Mostow (Corollary 3.7). The main tool in our proof of Theorem 3.2 is a geometric characterization of the semisimple elements of G (Lemma 3.6).

2. Preliminaries.

2.1 *Lie groups.* Given a Lie group G, G_0 will denote the identity component, $\mathfrak{G}$ will denote the Lie algebra, and $\exp: \mathfrak{G} \to G$ will denote the exponential map. G and $\mathfrak{G}$ are called *reductive* if the adjoint representation of $\mathfrak{G}$ (or, equivalently, of G_0) is fully reducible, i. e., if $\mathfrak{G}$ is the direct sum of an abelian ideal $\mathfrak{A}$ and a semisimple ideal $\mathfrak{G}'$; then $\mathfrak{A}$ is the center of $\mathfrak{G}$, $A = \exp(\mathfrak{A})$ is the identity component of the center of G_0 and is called the *connected center* of G_0, $\mathfrak{G}'$ is the derived algebra of $\mathfrak{G}$ and is called the *semisimple part* of $\mathfrak{G}$, $G' = \exp(\mathfrak{G}')$ is called the *semisimple part* of G_0, and there is a natural homomorphism $(a, g) \to ag$ of $A \times G'$ onto G_0.

If $\mathfrak{G}$ is reductive, then the Cartan subalgebras of $\mathfrak{G}$ are the subalgebras of the form $\mathfrak{A} \oplus \mathfrak{H}'$, where $\oplus$ denotes direct sum of ideals and $\mathfrak{H}'$ is a Cartan subalgebra of $\mathfrak{G}'$; thus the Cartan subalgebras of $\mathfrak{G}$ are abelian. By *Cartan subgroup* of a Lie group G, we mean a (necessarily connected) group of the form $\exp(\mathfrak{H})$ where $\mathfrak{H}$ is a Cartan subalgebra of $\mathfrak{G}$.

Under the adjoint representation of a Lie group G, an element $g \in G$ induces an automorphism $\mathrm{ad}(g)$ of $\mathfrak{G}$; we will call g *semisimple* if $\mathrm{ad}(g)$ is a fully reducible linear transformation of $\mathfrak{G}$. If g is a semisimple element of a reductive Lie group G, and if Z is the centralizer of g in G, then not only is Z reductive but the adjoint representation of G induces a fully reducible representation of $\mathfrak{Z}$ on $\mathfrak{G}$.

If S and T are subsets of a group G, then the commutator $[S, T]$ denotes the set of all elements $[s, t] = sts^{-1}t^{-1}$ where $s \in S$, $t \in T$.

If S and T are groups and ϕ is a homomorphism of S into the group of automorphisms of T, then the *semidirect product* $S \cdot_{\phi} T$ is the set $S \times T$ with the group structure $(s_1, t_1)(s_2, t_2) = (s_1 s_2, (\phi(s_2^{-1})(t_1)) t_2)$. If S is given as a group of automorphisms of T, then the semidirect product is denoted $S \cdot T$.

2.2. *Discrete groups.* A subgroup Γ of a topological group G is *discrete* if G has an open set U such that $\Gamma \cap U$ is just the identity element $1 \in G$. A subgroup H of G is *uniform* if the coset space $G/\bar{H}$ ($\bar{H}$ is the closure of H in G) is compact.

Let Γ be a discrete uniform subgroup of G. C. L. Siegel [9] has shown that G is locally compact, and, if every covering of G by open sets has a countable refinement,[8] then G has a compact subset F such that $\Gamma \cdot F = G$, every $g \in G$ has a neighborhood contained in a finite union of the γF, $\Gamma_F = \{\gamma \in \Gamma : \gamma F \text{ meets } F\}$ is finite, and Γ_F generates Γ if G is connected. It follows that Γ finitely generated if G/G_0 is finitely generated, but this is better seen directly [7]. F will be called a *fundamental domain* for the action of Γ on G by right translations.

2.3. *Symmetric spaces.* It is well known that a connected simply connected Riemannian symmetric space M is isometric to a product $M_0 \times M_1 \times \cdots \times M_t$ where M_0 is a Euclidean space and each M_i $(i > 0)$ is an irreducible Riemannian (non-Euclidean and not isometric to a product of lower dimensional Riemannian manifolds) symmetric space; M_0 is the *Euclidean part* of M, $M' = M_1 \times \cdots \times M_t$ is the *non-Euclidean part* of M, and the M_i $(i > 0)$ are the *irreducible factors* of M. We will say that M is *strictly non-Euclidean* if $M = M'$, i.e., if $\dim. M_0 = 0$, and will say that M is *strictly noncompact* if every irreducible factor of M is noncompact. If M is strictly noncompact, then every sectional curvature on M is $\leqq 0$.

Full groups of isometries are related by $I(M) = I(M_0) \times I(M')$, and $I(M')$ is generated by $I(M_1) \times \cdots \times I(M_t)$ together with all permutations on mutually isometric sets of M_i. Connected groups of isometries are related by $I_0(M) = I_0(M_0) \times I_0(M_1) \times \cdots \times I_0(M_t)$, and $I(M)/I_0(M)$ is finite. $I(M_0)$ is the Euclidean group $E(\dim. M_0)$, $I(M_i)$ is a compact semisimple Lie group if M_i is compact, and $I_0(M_i)$ is a noncompact centerless real or complex simple Lie group if M_i is noncompact $(i > 0)$.

Let S be a maximal connected flat (all sectional curvatures zero) totally geodesic submanifold of M. $I(M)$ acts transitively on the set of all such submanifolds, and the *rank* of M (denoted $\text{rank}. M$) is defined to be their common dimension. S is isometric to a product $M_0 \times S_1 \times \cdots \times S_t$ where S_i is a maximal connected flat totally geodesic submanifold of M_i, whence it is easily seen that S is a closed submanifold of M. The symmetry to M_i at a point of S_i induces a symmetry of S_i; it follows that each S_i (and thus S) is a connected Riemannian symmetric manifold of constant curvature zero. Thus S_i is a flat torus if M_i is compact and S_i is a Euclidean space if M_i ($i > 0$) is noncompact; to see this, we use [10, Théorème 4] or [11, § 14] together with the fact that S_i is the orbit of some Cartan subgroup of maximal vector rank in $I(M_i)$. If M_G is the product of the compact irreducible factors of M and M_N is the product of M_0 with the noncompact irreducible factors of M, it follows that S is isometric to the product of a flat torus of dimension $\text{rank}.\,M_G$ with a Euclidean space of dimension $\text{rank}.\,M_N$. For this reason, we define the *vector rank* of M (denoted $v\text{-rank}.\,M$) to be $\text{rank}.\,M_N$. Observe that $v\text{-rank}.\,M = \dim.\,M_0 + v\text{-rank}.\,M'$.

---

[8] Siegel requires that G have a countable basis for open sets, but uses only this weaker property.

Let Γ be a subgroup of $I(M)$. The action of Γ on M is *properly discontinuous* if every element of M has a neighborhood which meets its transforms by only a finite number of elements of Γ; this is equivalent to Γ being a discrete subgroup of $I(M)$. The action of Γ on M is *free* if $1 \neq \gamma \in \Gamma$ and $x \in M$ implies $\gamma(x) \neq x$. $M \to M/\Gamma$ is a covering space if and only if Γ acts freely and properly discontinuously on M. If M is strictly noncompact, then the isotropy subgroups of $I(M)$ are the maximal compact subgroups, and, if Γ acts properly discontinuously on M, it follows that Γ acts freely if and only if every element $\neq 1$ of Γ has infinite order. If Γ acts properly discontinuously on M, then M/Γ is a Hausdorff topological space (although it need not be a manifold), and M/Γ is compact if and only if Γ is a uniform subgroup of $I(M)$.

2.4. *Holonomy groups.* The *homogeneous holonomy group* $H(M,x)$ of a Riemannian manifold M at a point $x \in M$ is the group of linear transformations of the tangentspace M_x obtained by parallel translation of tangentvectors along sectionally smooth closed arcs based at x. The Riemannian metric gives M_x a positive definite inner product; $H(M,x)$ is a subgroup of the corresponding orthogonal group and carries the induced topology. The *restricted homogeneous holonomy group* is the identity component $H_0(M,x)$, consists of those elements of $H(M,x)$ obtained from nullhomotopic closed arcs, and is a closed subgroup of the orthogonal group of M_x; in particular, $H_0(M,x)$ is compact, and now $H(M,x)$ is compact if and only if it has only finitely many components. Thus we have a natural homomorphism of the fundamental group $\pi_1(M,x)$ onto the quotient $H(M,x)/H_0(M,x)$. If M is connected, then we speak of $H(M)$ and $H_0(M)$ in the same sense as $\pi_1(M)$.

Suppose that M is a connected simply connected Riemannian symmetric space, and $M = M_0 \times M_1 \times \cdots \times M_t$ is the decomposition into Euclidean and irreducible non-Euclidean parts. Then $H(M_0) = 1$ and $H(M_i, x)$ is the group of linear transformations of $(M_i)_x$ induced by the isotropy subgroup of $I_0(M_i)$ at x.

If M and N are Riemannian manifolds, $x \in M$ and $y \in N$, then

$$H(M \times N, (x,y)) = H(M,x) \times H(N,y).$$

3. Semisimplicity of discrete uniform subgroups.

3.1. If a Cartan subgroup of a semisimple Lie group G is normalized by an element $g \in G$, then it is known [5, Proposition 7.7] that g is a semisimple element of G. In order to make the estimates described in § 1, then, we need:

3.2. THEOREM.[4] *If* Γ *is a discrete uniform subgroup of a reductive Lie group* G *such that* G/G_0 *has no element of infinite order, then every element of* Γ *is a semisimple element of* G.

The essential part of the reduction to the semisimple case is given by:

3.3. LEMMA. *Let* Γ *be a discrete uniform subgroup of a connected reductive Lie group* G, *let* A *be the connected center of* G, *let* G' *be the semisimple part of* G, *and let* $\Gamma' = \{g' \in G' : g' = a\gamma \text{ for some } a \in A,\ \gamma \in \Gamma\}$. *Then* Γ' *is a discrete uniform subgroup of* G' *and* $\Gamma \cap A$ *is a discrete uniform subgroup of* A.

Proof. It is sufficient to consider the case where G' has no compact factor, for replacing G' by G'/K, where K is the maximal compact normal subgroup of G', affects neither hypotheses nor conclusions of the Lemma. Similarly, we may assume $G = A \times G'$.

Let $\{\gamma_i'\} \to 1$ be a sequence in Γ'; this gives us a sequence $\{\gamma_i\}$ in Γ with $\gamma_i = a_i\gamma_i'$, $a_i \in A$. A being central in G, $\{[\gamma_i,\gamma]\} = \{[\gamma_i',\gamma]\} \to 1$ for every $\gamma \in \Gamma$; thus γ_i commutes with γ for large i because Γ is discrete. Γ being finitely generated, it follows that γ_i is central in Γ for large i. Now let $\pi: G \to G'$ be the projection; $\Gamma' = \pi(\Gamma)$. Being uniform in G, Γ has the Selberg density property (S) in G ([8, Lemma 1] or [4, Lemma 1.4]); thus Γ' has the property (S) in G' [4, § 1.2]; it follows that the centralizer of Γ' in G' is just the center of G' [4, Corollary 4.4], whence γ_i' is central in G'

---

[4] As the proof will show, the essential case is that of a semisimple linear group, for which the result is known; see Borel and Harish-Chandra, "Arithmetic subgroups of algebraic groups," *Annals of Mathematics*, vol. 75 (1962), pp. 485-535, esp. § 11.2.

for large i. As G' has discrete center, this contradicts $\{\gamma_i'\} \to 1$. Thus Γ' is a discrete subgroup of G'.

Let F be a compact fundamental domain for the action of Γ on G by right translations. Γ' is a uniform subgroup of G' because $\pi(F)$ is compact and $G' = \Gamma' \cdot \pi(F)$.

Our proof that $\Gamma \cap A$ be uniform in A is a modification of an argument of A. Weil.[5] Let $a \in A$ and retain the notation above. Then $a = \gamma f$ for some $f \in F$ and some $\gamma \in \Gamma$ such that $\pi(\gamma) \in \pi(F)^{-1} \cap \pi(\Gamma)$. $\pi(F)^{-1}$ is compact and $\pi(\Gamma)$ was just seen discrete; this gives $\{\gamma_1, \cdots, \gamma_t\} \subset \Gamma$ such that the $\pi(\gamma_i)$ exhaust $\pi(F)^{-1} \cap \pi(\Gamma)$; it follows that $\gamma = \delta\gamma_i$ for some i and some $\delta \in \Gamma \cap A$. In other words, $a \in (\Gamma \cap A) \cdot \bigcup_{i=1}^{t} \gamma_i F$ for every $a \in A$. As $\bigcup_{i=1}^{t} \gamma_i F$ is compact, it follows that $A/(\Gamma \cap A)$ is compact. *q. e. d.*

The semisimple case is reduced to the linear semisimple case by means of:

3. 4. LEMMA. *Let* Γ *be a discrete uniform subgroup of a connected semisimple Lie group* G, *let* Z *be the center of* G, *and let* $\pi\colon G \to G/Z$ *be the projection. Then* $\pi(\Gamma)$ *is a discrete uniform subgroup of* G/Z *and* Γ *has finite index in* $\Gamma \cdot Z$.

Proof. It suffices to show $\Gamma \cdot Z$ discrete in G, and for this we may assume that G has no compact factor. Let $\{\gamma_i z_i\} \to 1$ be a sequence in $\Gamma \cdot Z$ with $\gamma_i \in \Gamma$ and $z_i \in Z$. Given $\gamma \in \Gamma$, $\{[\gamma_i, \gamma]\} = \{[\gamma_i z_i, \gamma]\} \to 1$; thus γ_i is central in Γ for large i because Γ is discrete and finitely generated. As in the previous lemma, it follows that $\gamma_i \in Z$ for large i, whence $\gamma_i z_i \in Z$ for i large. Z being discrete, this contradicts $\{\gamma_i z_i\} \to 1$. *q. e. d.*

3. 5. *Proof of Theorem* 3. 2. $\Gamma \cap G_0$ is a discrete uniform subgroup of G_0, G_0 satisfies the hypotheses of the Theorem, every element of Γ has a finite power in $\Gamma \cap G_0$, and γ^m cannot be semisimple unless γ is semisimple. Thus we may assume G connected. Semisimplicity of γ depending only on the automorphism $\mathrm{ad}(\gamma)$ of $\mathfrak{G}$, Lemmas 3. 3 and 3. 4 now allow us to assume that G is a centerless semisimple Lie group. Finally, every automorphism of a compact simple Lie algebra being semisimple (because it preserves the Killing form, which is negative definite), we may factor G by its maximal compact normal subgroup and assume that G has no compact factor. In summary, we need only consider the case where G is a product of noncompact centerless connected simple Lie groups.

---

[5] See his paper "Discrete subgroups of Lie groups II," *Annals of Mathematics*, vol. 75. Observer that the argument proves: If Δ is a discrete uniform subgroup of a connected group S and T is a closed normal subgroup of S, then $\Delta \cap T$ is uniform in T if and only if the image of Δ in S/T is discrete.

Let K be a maximal compact subgroup of G and let M be the Riemannian symmetric space G/K; G is the identity component of the full group of isometries of M, whence Γ is represented faithfully on M by isometries; the action of Γ on M is properly discontinuous because Γ is discrete in G. The adjoint representation of G represents Γ faithfully as a definitely generated real matrix group, so Γ has a subgroup of finite index with no element $\neq 1$ of finite order [8, Lemma 8]. We cut Γ down to this subgroup, and may thus assume that no element $\neq 1$ of Γ has a fixed point on M, so $M \rightarrow M/\Gamma$ is a covering of Riemannian manifolds which is a local isometry (a Riemannian covering). M/Γ is compact because G/Γ is compact. M being a complete connected simply connected Riemannian manifold with every sectional curvature $\leqq 0$, it follows [6] that every $\gamma \in \Gamma$ preserves some geodesic on M. Theorem 3.2 follows from:

3.6. LEMMA. *Let M be a strictly non-Euclidean Riemannian symmetric space, let $G = I(M)$, the full group of isometries of M, and let $g \in G$. Then g is a semisimple element of G if and only if some power g^m, $m \neq 0$, preserves a geodesic on M.. If g preserves a geodesic σ through $x \in M$ and has no fixed point on σ, then $g = kp = pk$ where $k \in G$, $k(x) = x$, and p is a transvection along σ.*

Proof. Let g be semisimple. Then g normalizes a Cartan subalgebra $\mathfrak{H}$ of $\mathfrak{G}$ [5, Theorem 7.6]. Choose $m \geqq 1$ such that $g^m \in H = \exp(\mathfrak{H})$; this is possible because G is semisimple and G/G_0 is finite. There is an element $x \in M$ and a Cartan decomposition $\mathfrak{G} = \mathfrak{K} + \mathfrak{P}$ where $\mathfrak{K}$ is the Lie algebra of the isotropy subgroup K of G at x, such that $\mathfrak{H} = (\mathfrak{H} \cap \mathfrak{K}) + (\mathfrak{H} \cap \mathfrak{P})$. This holds for compact G_0 by conjugacy of maximal tori and because an involutive automorphism must conserve a maximal torus; it is known for noncompact linear simple G_0 ([12], p. 107); it now follows in our case. Now $g^m = kp$ with $k \in H \cap K$ and $p = \exp(X)$ for some $X \in \mathfrak{H} \cap \mathfrak{P}$; thus g^m preserves the geodesic $\sigma = \{\exp(tX)x\}$ on M.

Let g^m preserve a geodesic σ on M; we wish to show g semisimple, and it suffices to show g^m semisimple. Thus we may assume g to preserve σ. As g^2 preserves σ, we now replace g by g^2 if g has a fixed point on σ. σ being a totally geodesic submanifold of M, g induces an isometry of σ onto itself, and the possible replacement of g by g^2 shows that $g\colon \sigma_t \rightarrow \sigma_{t+a}$ for some real number a, where t is arc length. Let $x \in \sigma$, say $x = \sigma_0$, and take the Cartan decomposition $\mathfrak{G} = \mathfrak{K} + \mathfrak{P}$ at x; this gives $X \in \mathfrak{P}$ with $\sigma_t = \exp(tX)x$. $k = \exp(-aX)g \in K$, and k commutes with X because it preserves every σ_t, whence k commutes with $p = \exp(aX)$. Now $g = kp = pk$, k is semisimple because it lies in a compact group K, p is a transvection along σ, and p is semisimple because it is represented on $\mathfrak{G}$ by a positive definite matrix. Being the product of two commuting semisimple elements, g is semisimple. *q. e. d.*

3.7. COROLLARY. *If Γ is a discrete uniform subgroup of a reductive Lie group G such that G/G_0 has no element of infinite order, and if Δ is an abelian subgroup of Γ, then Δ normalizes a Cartan subgroup of G.*

Proof. Theorem 3.2 says that $\mathrm{ad}(\Delta)$ is an abelian group of semisimple automorphisms of $\mathfrak{G}$; thus $\mathrm{ad}(\Delta)$ is diagonable on $\mathfrak{G}$ over the complex numbers, and it follows that $\mathrm{ad}(\Delta)$ has a finitely generated subgroup D such that every D-invariant subspace of $\mathfrak{G}$ is $\mathrm{ad}(\Delta)$-invariant. Finitely generated and abelian, D is of type $(MP)^{*}$ (see [5, p. 404]); thus D leaves invariant a Cartan subalgebra $\mathfrak{H}$ of $\mathfrak{G}$ [5, Theorem 7.6]. $\mathfrak{H}$ is $\mathrm{ad}(\Delta)$-invariant by choice of D, whence Δ normalizes the Cartan subgroup $\exp(\mathfrak{H})$ of G. *q. e. d.*

3.8. *Remark.* If Ψ is a subgroup of Γ which has normal subgroups Ψ_i such that $\Psi = \Psi_k \supset \Psi_{k-1} \supset \cdots \supset \Psi_1 \supset \Psi_0 = 1$ with Ψ_i/Ψ_{i-1} cyclic, then $\mathrm{ad}(\Psi)$ is a group of semisimple automorphisms of type $(MP)^{*}$ of $\mathfrak{G}$, so Ψ normalizes a Cartan subgroup of G by [5, Theorem 7.6].

4. Bounds for reductive groups.

4.1. Let G be a reductive Lie group. G has only a finite number of conjugacy classes of Cartan subgroups; let $\{H_1, \cdots, H_m\}$ be a maximal collection of mutually nonconjugate Cartan subgroups of G. H_i being a connected abelian Lie group of dimension $r = \mathrm{rank}.\, G$, it is isomorphic to the product of a vector group R^{u_i} with a torus T^{r-u_i}. The *vector rank* of G, denoted v-rank. G, is defined to be the maximum of the u_i. If G/G_0 is finite, then each H_i has finite index in its normalizer in G; in this case there is a smallest integer, which we define to be the *torsion rank* of G and denote t-rank. G such that every finite abelian subgroup of G (which will automatically normalize a Cartan subgroup by [5, Theorem 7.6]) can be expressed as the product of $\leqq t$-rank. G cyclic groups.

4.2. THEOREM. *Let Γ be a discrete uniform subgroup of a reductive Lie group G such that G/G_0 is finite, and let Δ be an abelian subgroup of Γ. Then Δ can be expressed as the product of $\leqq t$-rank. G finite cyclic groups with a free abelian group on $\leqq v$-rank. G generators; in particular, Δ is finitely generated. Furthermore, the second bound is best possible in the sense that Γ has a subgroup which is free abelian on v-rank. G generators. Finally, if Σ is a subgroup of Γ which is the product of a finite abelian group and a* *free abelian group on m generators, if H is a Cartan subgroup of G normalized by Σ, and if Γ does not have an abelian subgroup Ψ such that $\Sigma \cap \Psi$ be of finite index in Σ but of infinite index in Ψ, then $\Sigma \cap H$ is uniform in H.*

Proof. By Corollary 3.7, Δ normalizes a Cartan subgroup A of G. $\Delta \cap A$ is finitely generated because it is a discrete subgroup of the connected abelian Lie group A, and $\Delta \cap A$ has finite index in Δ because A has finite index in its normalizer in G; thus Δ is finitely generated. The first statement now follows from the structure theorem for finitely generated abelian groups and the definitions of t-rank. G and v-rank. G. We need some lemmas for the other statements.

4.3. LEMMA.[6] *Let Γ be a discrete uniform subgroup of a Hausdorff topological group G, let D be a finitely generated subgroup of Γ, and let G_D and Γ_D be the respective centralizers of D in G and Γ. Then Γ_D is a discrete uniform subgroup of G_D.*

Proof. Let $\pi\colon G \to G/\Gamma$ be the projection. As $\Gamma_D = \Gamma \cap G_D$ and G/Γ is a compact Hausdorff space, Γ_D is uniform in G_D if and only if $\pi(G_D)$ is closed in G/Γ. If $\pi(G_D)$ is not closed in G/Γ, then $\pi^{-1}\pi(G_D)$ is not closed in G, so there is a sequence $\{\gamma_i\}$ in Γ of elements distinct mod G_D and a sequence $\{g_i\}$ in G_D with $\{\gamma_i g_i\} \to x$ for some $x \in G$. Given $d \in D$, $\{[\gamma_i, d]\} = \{[\gamma_i g_i, d]\} \to [x, d]$, whence $[\gamma_i, d] = [x, d]$ for large i because Γ is discrete, implying that $\gamma_i^{-1}\gamma_j$ commute with d for large i and j. As D is finitely generated, $\gamma_i^{-1}\gamma_j \in G_D$ for large i and j, contradicting our choice of the sequence $\{\gamma_i\}$. *q.e.d.*

4.4. We will prove the last statement of the theorem, retaining the notation of Lemma 4.3. Theorem 3.2 and an induction on the number of generators of Σ show that G_Σ is a reductive Lie group, so its Lie algebra is a direct sum $\mathfrak{A} \oplus \mathfrak{C} \oplus \mathfrak{N}$ with $\mathfrak{A}$ abelian, $\mathfrak{C}$ a sum of compact simple Lie algebras, and $\mathfrak{N}$ a sum of noncompact simple Lie algebras. By Lemma 3.3 and assumption on Σ, we see that $\mathfrak{N} = 0$; thus $(G_\Sigma)_0 = V \times K$ where V is a vector group in $\exp(\mathfrak{A})$ and K is a compact group containing $\exp(\mathfrak{C})$. $\Sigma \cap H$ having finite index in Σ, and $H \subset (G_\Sigma)_0$, gives $m \leqq v$-rank. $H \leqq$ dim. V. On the other hand, $V \cap \Gamma_\Sigma$ is free abelian on dim. V generators by Lemmas 4.3 and 3.3; thus dim. $V \leqq m$ by assumption on Σ. It follows that $m =$ v-rank. H, proving the last statement of the Theorem.

4.5. LEMMA. *Let H be a Cartan subgroup of a reductive Lie group G,*

---

[6] This is essentially the same as A. Selberg's result [8, Lemma 2].

and suppose that H *has an element* h *such that, given* $g \in G_0$, *the number of distinct absolute values among the eigenvalues of* $\mathrm{ad}(h)$ (*acting on* $\mathfrak{G}$) *is at least as large as the number of distinct absolute values among the eigenvalues of* $\mathrm{ad}(g)$. *Then* $v\text{-rank}.\, H = v\text{-rank}.\, G$.

For $\mathfrak{H}$ has the maximal number of linearly independent real-valued roots among all Cartan subalgebras of $\mathfrak{G}$.

4.6.[7] We will finish the proof of Theorem 4.2 by proving the second statement. Let $g \in G_0$, the number of whose absolute values of eigenvalues in the adjoint representation is maximal among the elements of G_0. Taking powers only separates further the absolute values of the eigenvalues, so we have a neighborhood U of 1 in G_0 and an integer m such that:

1. If μ is the Haar measure on G/Γ and $\pi\colon G \to G/\Gamma$ is the projection, then $m > \mu(G/\Gamma)/\mu(\pi(U))$.
2. If $1 \leqq a \leqq m$ and $u_i \in U$, then the number of distinct absolute values of eigenvalues of $\mathrm{ad}(u_1 g^a u_2^{-1})$ is not less than the number for $\mathrm{ad}(g)$.

We choose [8, Lemma 1] the integer a and the $u_i \in U$ such that $u_1 g^a u_2^{-1} = \gamma \in \Gamma$, replace γ by a power which lies in a Cartan subgroup H of G, and observe that $v\text{-rank}.\, H = v\text{-rank}.\, G$ by Lemma 4.5. We may take γ to be a regular element of G, whence $H/(H \cap \Gamma)$ is compact by Lemma 4.3 with $D = \{\gamma\}$. Thus $H \cap \Gamma$ has a subgroup which is free abelian on v rank. G generators. The Theorem is proved. *q.e.d.*

5. Bounds for groups with Euclidean factor.

Our tool for the treatment of groups with Euclidean factor is the following generalization of theorems of L. Bieberbach [3] and L. Auslander [1]:

5.1. THEOREM. *Let* K *be a compact group of automorphisms of a connected simply connected nilpotent Lie group* N, *let* E *be the semidirect product* $K \cdot N$, *let* $L = E \times G$ *where* G *is a reductive Lie group with* G/G_0 *finite, and let* Γ *be a discrete uniform subgroup of* L. *Then* $\Gamma \cap (N \times G)$ *is a normal subgroup of finite index in* Γ *which is a discrete uniform subgroup of* $N \times G$.

Proof.[8] *As* in § 3.5, we may assume G to be connected and without compact normal subgroup; then $G = A \cdot G'$ where A, the connected center of G, is a vector group and G', the semisimple part of G, is without compact factor. Let $p: A \times G' \to G$ be the projection; then $(1 \times p)^{-1}(\Gamma)$ is a discrete uniform subgroup of $E \times A \times G'$ with the same projection on K as has Γ. Replacing N by $N \times A$, we see that it suffices to prove the Theorem when G is semisimple and without compact factor.

---

[7] Compare with paragraph 3, p. 151, of [8], where A. Selberg considers the case $G = SL(n; R)$.

[8] This Theorem can be proved as a consequence of [2, Theorem 1], to which it is similar. We find it more convenient, however, to give an argument which is a variation on the proof of that result.

Let J be the closure of $\Gamma \cdot N$ in L, and let $\alpha: L \to G$ and $\beta: L \to K$ be the projections. J_0 is solvable by the generalized Zassenhaus Lemma [2, Proposition 2], and J is clearly normalized by Γ; as $\alpha(\Gamma)$ has the Selberg density property (S) in G (for Γ has it in L), $\alpha(\Gamma)$normalizes $\alpha(J)$ and G is semisimple without compact factor, it follows [4, Theorem 4.1] that $\alpha(J)$ has discrete closure in G. Thus $\alpha(\Gamma)$ is discrete, being contained in $\alpha(J)$ The proof of the last part of Lemma 3.3 now shows $\Gamma \cap E$ uniform in E.

$\Gamma \cap E$ being a discrete uniform subgroup of E, the generalized Bieberbach Theorem [1, Theorem 1] shows that $(\Gamma \cap E) \cap N = \Gamma \cap N$ is a discrete uniform subgroup of N. $\Gamma \cap N$ is normal in Γ, and is thus normalized by $\beta(\Gamma)$. On the other hand, interpreting K as a group of automorphisms of N, uniformity of $\Gamma \cap N$ in N implies that an element of K is determined by its action on $\Gamma \cap N$ [1, Theorem 2]. It follows that $\beta(\Gamma)$ is finite. *q. e. d.*

5.2. COROLLARY. *Let Γ be a discrete uniform subgroup of $L = E \times G$ where E is a semidirect product $K \cdot V$, K is a compact group of automorphisms of the vector group V, and G is a reductive Lie group with G/G_0 finite. If Δ is an abelian subgroup of Γ, then Δ normalizes some Cartan subgroup H of G, and Δ can be expressed as the product of $\leqq$ t-rank.$(K \times G)$ finite cyclic groups with a free abelian group on $\leqq$ dim. $V +$ v-rank. G generators. If Γ has no abelian subgroup Σ with the property that $\Delta \cap \Sigma$ has finite index in Δ and infinite index in Σ, then $\Delta \cap (V \times H)$ is uniform in $V \times H$. Finally, Γ has a subgroup which is free abelian on dim. $V +$ v-rank. G generators.*

Proof. By Theorems 3.2 and 5.1, every element of Γ is a semisimple element of the reductive Lie group $G'' = \Gamma \cdot (V \times G)$, and G''/G_0'' is finite. The Cartan subgroups of G'' being of the form $V \times$(Cartan subgroup of G), we have v-rank. $G'' =$ dim. $V + v$-rank. G. Every finite subgroup of E being conjugate to a subgroup of K, we have t-rank. $G'' \leqq t$-rank. $(K \times G)$. The Corollary now follows from Theorem 4.2. *q. e. d.*

6. Application to symmetric spaces.

6.1. Let M be a connected simply connected Riemannian symmetric space, and let $L = I(M)$, the full group of isometries of M. If $M = M_0 \times M'$

is the decomposition of M into the product of its Euclidean and non-Euclidean parts, then $L = E \times G$ where $G = I(M')$ is a semisimple (and thus reductive) Lie group with G/G_0 finite and $E = I(M_0)$ is the Euclidean group $E(n)$, $n = \dim. M_0$, semidirect product $O(n) \cdot V$ where $O(n)$ is the orthogonal group of a vectorspace V which can be identified with M_0. As every flat maximal connected totally geodesic submanifold of M' is an orbit of a Cartan subgroup of maximal vector rank in G, we have v-rank. $M' = v$-rank. G. Thus v-rank. $M = \dim. M_0 + v$-rank. $I(M')$.

6.2. THEOREM. *Let M be a connected simply connected Riemannian symmetric space, let M_0 and M' be the Euclidean and non-Euclidean parts of M, let Γ be a properly discontinuous group of isometries of M with M/Γ compact, and let Δ be an abelian subgroup of Γ. Then M has a closed connected Δ-invariant flat totally geodesic submanifold S_Δ whose image in M/Γ is compact, and Δ can be expressed as the product of* $\leqq t$-rank. $(O(\dim. M_0) \times I(M'))$ *finite cyclic groups with a free abelian group on* $\leqq v$-rank. M *generators. If Γ has no abelian subgroup Σ with the property that $\Delta \cap \Sigma$ has finite index in Δ but infinite index in Σ, then S_Δ/Δ is compact. Finally, Γ has a subgroup Ψ which is free abelian on* v-rank. M *generators, and S_Ψ can be taken to be a maximal connected flat totally geodesic submanifold of M.*

Proof. Γ acts by isometries, M/Γ is compact, and $I(M)$ acts transitively on M with compact isotropy subgroups; it follows that Γ is a discrete uniform subgroup of $I(M)$.

6.3. Retaining the notation of § 6.1, we will define S_Δ to be an orbit of one of the groups $V \times H$ where H is a Cartan subgroup of $I(M')$ normalized by Δ; such groups exist by Corollary 5.2. When the choice of H is made, we will choose $x \in M_0$ to be the origin of V and choose $y \in M'$ such that $\mathfrak{H} = (\mathfrak{H} \cap \mathfrak{K}) + (\mathfrak{H} \cap \mathfrak{P})$ where K is the isotropy subgroup of $I(M')$ at y and $\mathfrak{P}$ is the orthogonal complement of $\mathfrak{K}$ in $\mathfrak{I}(M')$ under the Killing form of $\mathfrak{I}(M')$. Then $S_\Delta = (V \times H)(x,y) = M_0 \times H(y)$ is clearly closed, connected, flat and totally geodesic in M. Let $\pi\colon I(M) \rightarrow I(M')$ be the projection and let $\delta \in \Delta$. $\pi(\Delta)$ normalizes H, and is thus contained in $K \cdot H$; now $\pi(\delta) = kh$. This gives $\pi(\delta)H(y) = khH(y) = kHk^{-1}(y) = H(y)$, proving S_Δ to be Δ-invariant.

We choose H to be a Cartan subgroup of $I(M')$ which is normalized by some maximal abelian subgroup Φ of Γ which contains Δ, and set $\Phi' = \Phi \cap (V \times H)$; Φ' has finite index in Φ by Theorem 5.1 and finiteness of G/G_0. Φ is finitely generated, and thus, by Lemma 4.3, is uniform in its centralizer A in $I(M)$; it follows that Φ' is uniform in $V \times H$. Thus the image of S_Δ in M/Γ is compact. As the bounds on the size of Δ follow from Corollary 5.2, this proves the first statement of the Theorem.

6.4. Observe that not only is the image of S_Δ in M/Γ compact, but S_Δ/Φ is compact. Thus S_Δ/Δ is compact if Δ has finite index in Φ. The structure of abelian subgroups of Γ is such that Δ has finite index in Φ if Γ has no abelian subgroup Σ with the property that $\Delta \cap \Sigma$ has finite index in Δ but infinite index in Σ. This proves the second statement of the Theorem.

Corollary 5.2 shows that Γ has a subgroup which is free abelian on v-rank.M generators. If Δ is such a subgroup, then H is a Cartan subgroup of maximal vector rank in $I(M')$. Examining the compact and noncompact factors of M' separately, we see that, for proper choice of $y \in M'$, S_Δ is a maximal connected flat totally geodesic submanifold of M. *q. e. d.*

6.5. The complete connected locally symmetric Riemannian manifolds are precisely those manifolds whose universal Riemannian covering manifold is symmetric. Thus Theorem 6.2 gives us:

6.6. COROLLARY. *Let N be a compact connected locally symmetric Riemannian manifold, let D be an abelian subgroup of the fundamental group $\pi_1(N)$, and let M_c be the product of the compact irreducible factors of the universal Riemannian covering manifold M of N. Then N has a closed connected flat totally geodesic submanifold S_D and an element $x \in S_D$ such that D is represented by closed geodesic arcs in S_D based at x, and D can be expressed as the product of $\leqq t$-rank.$I(M_c)$ finite cyclic groups with a free abelian group on $\leqq v$-rank.M generators; thus D is free abelian on $\leqq v$-rank.M generators if M is strictly noncompact. $\pi_1(N)$ has a subgroup P which is free abelian on v-rank.M generators, and S_P can be taken to be a maximal connected flat totally geodesic submanifold of N; thus N has a maximal connected flat totally geodesic submanifold which is closed in N.*

7. Holonomy groups of locally symmetric spaces. Corollary 7.2 is the only part of § 7 which uses the results of preceding sections; in fact, it uses only Theorem 5.1, which does not depend on preceding results.

7.1. THEOREM. *Let Γ be the group of deck transformations of the universal Riemannian covering $\pi\colon M \rightarrow N$ of a complete locally symmetric Riemannian manifold N, let $M = M_0 \times M'$ be the decomposition into Euclidean and non-Euclidean parts, and let V be the group of pure translations of M_0. Then there is a canonical isomorphism between*

$$\Gamma \cdot (V \times I_0(M'))/(V \times I_0(M'))$$

and the group $H(N)/H_0(N)$ *of components of the homogeneous holonomy group of* N.

Proof. Recall the homomorphism β of Γ onto $H(N,\pi(x))/H_0(N,\pi(x))$ defined by $\beta(\gamma)=t_\gamma\cdot H_0(N,\pi(x))$, where t_γ is the operation defined by $\pi(\tau_\gamma)$ and τ_γ is any sectionally smooth arc in M from x to $\gamma(x)$. We can represent t_γ on the tangentspace M_x as the differential $\gamma_*: M_x\rightarrow M_{\gamma(x)}$ followed by parallel translation of tangentvectors backwards along τ_γ.

Let $x\in M=M_0\times M'$ have repesentation $x=(x_0,x')$, let K' be the isotropy subgroup of $I(M')$ at x', and let $P'=\exp(\mathfrak{P}')$ where $\mathfrak{P}'$ is the orthogonal complement of $\mathfrak{K}'$ in $\mathfrak{I}(M')$ under the Killing form of $\mathfrak{I}(M')$. The every element of $I(M')$ has expression $p'k'$ with $k'\in K'$ and $p'\in P'$. Observe that the identity component K_0' is the isotropy subgroup of $I_0(M')$ at x', and its action on the tangentspace M_x is that of $H(M,x)=H_0(N,\pi(x))$; also, P' is the set of transvections along geodesics in M' which pass through x', and $p_*': M_{x'}'\rightarrow M_{p'(x')}'$ is parallel translation along the geodesic arc from x' to $p'(x')$ on which p' is a transvection.

Let $\gamma\in\Gamma$; $\gamma=\gamma_0\gamma'$ with $\gamma_0\in I(M_0)$ and $\gamma'\in I(M')$. $\gamma'=p'k'$ as above and $\gamma_0=p_0k_0$ with $k_0(x_0)=x_0$ and $p_0\in V$; thus $\gamma=(p_0p')(k_0k')$. Let $y=(y_0,y')$ be the image of x; then p_0 is transvection along a geodesic arc τ_0 in M_0 from x_0 to y_0, p' is transvection along a geodesic arc τ' in M' from x' to y', we define τ_γ to be the geodesic arc in M from x to y with projections τ' and τ_0 and it is then clear that t_γ is represented by the differential of k_0k' on M_x. Thus the canonical homomorphism β of Γ onto $H(N)/H_0(N)$ induces the isomorphism of the Theorem. *q. e. d.*

7.2. COROLLARY. *Let* N *be a compact locally symmetric Riemannian manifold. Then the homogeneous holonomy group* $H(N)$ *is compact, i. e.,* $H(N)/H_0(N)$ *is finite.*

This follows easily from Theorems 5.1 and 7.1.

7.3. COROLLARY. *Let* M *be a connected simply connected Riemannian symmetric space with Euclidean part* M_0 *and non-Euclidean part* M'. *If* $\dim. M_0>2$, *or if* $\dim. M_0=2$ *and* M' *is noncompact, then there are continuum many affinely inequivalent diffeomorphic Riemannian manifolds covered by* M *which have noncompact homogeneous holonomy groups. If* $\dim. M_0<2$, *or if* $\dim. M_0=2$ *and* M' *is compact, then every Riemannian manifold covered by* M *has compact homogeneous holonomy group. If* $\dim. M_0=q-1<2$ *and* r *is the order of* $I(M')/I_0(M')$, *then the number of components of the homogeneous holonomy group of a Riemannian manifold covered by* M *is a divisor of* qr.

Proof.[9] The last statement follows from Theorem 7.1 and the fact the group of translations of M_0 has finite index q in $I(M_0)$. This also proves the second statement except when dim. $M_0 = 2$. Let dim. $M_0 = 2$, let M' be compact, and let Γ be the group of deck transformations of a Riemannian covering $M \to N$. $I(M')$ is compact, whence the projection Γ_0 of Γ on $I(M_0)$ is discrete. We wish to show that the group H of rotation parts of elements of Γ_0 is finite. Let U be the linear subspace of the vectorspace M_0 which is spanned by the translation parts of elements of Γ_0. If dim. $U = 2$, then finiteness of H follows from the Bieberbach theorem [3] (or from Theorem 5.1). If dim. $U = 1$, then M_0 has an orthonormal basis $\{u, v\}$ where u spans U. As H normalizes U, every element of H has matrix $\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$ in this basis; thus H is finite. If dim. $U = 0$, then $\Gamma_0 = H$ lies in a compact group, and thus finite because Γ_0 is discrete. Now H is finite in any case, and the last part of the second statement follows from Theorem 7.1.

7.4. For each real number t, we define $g_t = \begin{pmatrix} \cos(2\pi t) & \sin(2\pi t) \\ -\sin(2\pi t) & \cos(2\pi t) \end{pmatrix}$, the rotation with eigenvalues $\exp(\pm 2\pi\sqrt{-1}t)$. Now suppose dim. $M_0 \geqq 2$, view M_0 as a vectorspace, and let $\{v_1, \cdots, v_r\}$ be an orthonormal basis of M_0; let A_t be the linear transformation $\begin{pmatrix} g_t & 0 \\ 0 & I_{r-2} \end{pmatrix}$ of M_0. If dim. $M_0 > 2$, then let γ_t be the isometry $(m_0, m') \to (A_t m_0 + v_3, m')$ of $M = M_0 \times M'$; if dim. $M_0 = 2$ and M' is noncompact, then we have a tranvection τ in a noncompact irreducible factor of M', and we define γ_t to be the isometry $(m_0, m') \to (A_t m_0, \tau m')$ of M. In either case, γ_t generates an infinite cyclic subgroup Γ_t of $I(M)$ which acts freely and properly discontinuously on M. Thus $N_t = M/\Gamma_t$ is a Riemannian manifold covered by M. In both cases $\gamma_t = \beta_t \alpha_t$ where β_t is a transvection of M along some geodesic σ through our basepoint x, and where α_t is an isometry of M with $\alpha_t(x) = x$. The element of $H(N_t)/H_0(N_t)$ determined by γ_t being represented on the tangentspace M_x by the differential of γ_t follows by parallel translation along σ from $\beta_t(x) = \gamma_t(x)$ to x, this element is represented on M_x by the differential of α_t. By construction, this differential is given by A_t on $(M_0)_x$ and is the identity on $(M')_x$; thus we may view the linear transformation A_t as a generator of $H(N_t)/H_0(N_t)$. In particular, N_t has compact homogeneous holonomy group if and only if A_t has finite order, i.e., if and only if t is rational. Affine equivalence induces isomorphism of holonomy groups as groups of linear transformations; thus the first statement follows from the fact that the N_t are mutually real-analytically homeomorphic and we can choose continuum many algebraically independent irrational numbers t.

---

[9] As the proof will show, the essential case in when M is irreducible. This was originally handled by a lemma developed in discussions with H. C. Wang; in the present context, however, it is easier to appeal to Theorem 7.1.

q. e. d.

THE UNIVERSITY OF CALIFORNIA AT BERKELEY,
THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] L. Auslander, "Bieberbach's theorems on space groups and discrete uniform subgroups of Lie groups," *Annals of Mathematics*, vol. 71 (1960), pp. 579-590.

[2] ———, "On radicals of discrete subgroups of Lie groups with application to locally affine spaces," *American Journal of Mathematics*, to appear.

[3] L. Bieberbach, "Über die Bewegungsgruppen der Euklidischen Räume I," *Mathematische Annalen*, vol. 70 (1911), pp. 297-336; II, *Mathematische Annalen*, vol. 72 (1912), pp. 400-412.

[4] A. Borel, "Density properties for certain subgroups of semisimple groups without compact components," *Annals of Mathematics*, vol. 72 (1960), 179-188.

[5] A. Borel and G. D. Mostow, "On semi-simple automorphisms of Lie algebras," *Annals of Mathematics*, vol. 61 (1944), 389-405.

[6] É. Cartan, *Lecons sur la géométrie des espaces de Riemann*, Note III, 2ieme éd., Gauthier Villars, Paris, 1946.

[7] A. M. Macbeath and S. Świerczkowski, "On the set of generators of a subgroup," *Koninklijke Nederlandse Akademie van Wetenschappen. Indagationes Mathematicae ex Actis Quibus Titulus. Proceedings of the Section of Science.* (Amsterdam), vol. 21 (1959), pp. 280-281.

[8] A. Selberg, "On discontinuous groups in higher-dimensional symmetric spaces," *Contributions to Function Theory*, Tata Institute, Bombay, 1960, pp. 147-164.

[9] C. L. Siegel, "Discontinuous groups," *Annals of Mathematics*, vol. 44 (1943), pp. 674-689.

[10] J. A. Wolf, "Sur la classification des variétés riemanniennes homogènes à courbure constante," *Comptes Rendus des Seances de l'Academie des Sciences* (Paris), vol. 250 (1960), pp. 3443-3445.

[11] ———, "Homogeneous manifolds of constant curvature," *Commentarii Mathematici Helvetici*, vol. 36 (1961), pp. 112-147.

[12] Harish-Chandra, "The characters of semi-simple Lie groups," *Transactions of the American Mathematical Society*, vol. 83 (1956), pp. 98-163.

PERTURBATIONS OF SELF-ADJOINT TRANSFORMATIONS.*

By Louis de Branges.

An interesting relationship exists between two self-adjoint transformations S and T in a Hilbert space $\mathcal{H}$ if they are close to each other in a suitable sense. Although the point spectrum of a transformation is sensitive to small changes, the absolutely continuous part of the spectrum has stability properties—a fact which seems to have first been recognized by Friedrichs [6]. In support of this principle, Rosenblum [12] has shown that the transformations S and T are unitarily equivalent if they have absolutely continuous spectral measures and if the difference $T-S$ is of trace class. Then Kato [8] showed that the absolutely continuous parts of S and T are unitarily equivalent in any case if $T-S$ is of trace class. Finally, Kuroda [9, 10] obtained the same conclusion under the weaker hypothesis that $(T-w)^{-1}-(S-w)^{-1}$ is of trace class for some, and hence every, non-real number w. This condition will be referred to as the Kuroda hypothesis.

These theorems have a curious proof which depends on ideas from the scattering problem of quantum mechanics, discussed from this point of view by Kuroda [11]. The unitary transformation which achieves the desired unitary equivalence is obtained thereby as a limit of commutators of operators constructed from S and T. When $T-S$ has one-dimensional range, Kato [7] has given a more explicit description of the unitary transformation. This computation has led Rosenblum [13] to ask whether, in general, the limit of commutators can be evaluated so as to exhibit the unitary transformation. In this paper we construct explicitly a unitary transformation which establishes the desired equivalence, but we do not know that it is the same as the one resulting from the quantum mechanical formalism.

Acknowledgement is made to Marvin Rosenblum for discussion of perturbation theory (Institute for Advanced Study, 1959-1960), which stimulated the present work. Our approach is also influenced by work [1-4] with entire functions, and in particular by a generalization of the Hilbert transformation which occurs in [3]. The present use of vector valued analytic functions is derived from [5]. To describe our results we must first introduce vector notation.

Let $\mathcal{C}$ be a fixed, separable Hilbert space. By a vector we always mean

---

* Received March 14, 1962; revised September 24, 1962.

an element of this space. A special notation is used for the norm and inner product of vectors to avoid confusion with other spaces we shall construct. Let $|c|$ be the norm of a vector c. If b is a vector, let $\bar{b}$ be the corresponding linear functional on vectors so that the inner product takes the form $\langle a, b\rangle = \bar{b}a$. By an operator we mean a bounded linear transformation of vectors into vectors. If A is an operator, let $|A|$ be the operator norm and let $\bar{A}$ be the adjoint of A. We write $A \geqq 0$ if $\bar{c}Ac \geqq 0$ holds for every vector c, with the implication that $A = \bar{A}$ is self-adjoint. If A and B are operators, the inequality $A \leqq B$ means that $B - A \geqq 0$. An operator is said to be invertible if it has everywhere defined and bounded inverse A^{-1}. If a and b are vectors, let $a\bar{b}$ be the corresponding operator defined by $(a\bar{b})c = a(\bar{b}c)$ for every vector c. The theory requires the choice of a fixed invertible operator I such that $\bar{I} = -I = I^{-1}$. Let $\mathcal{C}_+$ be the kernel of $I - i$ and let $\mathcal{C}_-$ be the kernel of $I + i$. Then, $\mathcal{C}_-$ and $\mathcal{C}_+$ are orthogonal subspaces of $\mathcal{C}$ which span $\mathcal{C}$. We use τ for the finite or infinite trace norm of an operator, and σ for the Schmidt norm.

A vector valued function $f(z)$, defined in a region Ω of the complex plane, is said to be analytic in the region if the complex valued function $\bar{c}f(z)$ is analytic in the region for every choice of vector c. An operator valued function $F(z)$, defined in Ω, is said to be analytic in the region of the complex valued function $\bar{b}F(z)a$ is analytic in the region for every choice of vectors a and b.

We use also an integration theory for operator valued measures. By a non-negative, operator valued measure μ, we mean a countably additive function, defined for bounded Borel subsets of the real line, whose values are non-negative operators. By countably additive we mean that $\mu(\cup E_n) = \sum \mu(E_n)$ holds in the weak sense whenever (E_n) is a disjoint sequence of Borel sets with a bounded union. Corresponding to μ, we construct a Hilbert space $L^2(\mu)$, a dense vector subspace of which consists of elements which are equivalence classes of vector valued functions. Perhaps through ignorance we find it necessary to suppose that the value of μ are of trace class for a meaningful construction. Let $f = c_1\chi E_1 + \cdots + c_r\chi E_r$ be a simple vector valued function, where $\chi E_1, \cdots, \chi E_r$ are the characteristic functions of disjoint bounded Borel sets $E_1, \cdots, E_r$, and where $c_1 \cdots, c_r$ are corresponding vectors. The definition

$$\int \bar{f}(t)\, d\mu(t) f(t) = \bar{c}_1\mu(E_1)c_1 + \cdots + \bar{c}_r\mu(E_r)c_r$$

is seen not to depend on the choice of representation by characteristic functions, and

$$0 \leqq \int \bar{f}(t)\, d\mu(t) f(t) \leqq \int |f(t)|^2 d\tau(\mu(t)).$$

More generally, if $f(x)$ is a Borel measurable, vector valued function and if $\int |f(t)|^2 d\tau(\mu(t))$ is finite, define

$$\int \bar{f}(t)\,d\mu(t) f(t) = \lim \int \bar{f}_n(t)\,d\mu(t) f_n(t),$$

where $(f_n(x))$ is a sequence of simple, vector valued functions such that

$$\lim \int |f(t) - f_n(t)|^2 d\tau(\mu(t)) = 0.$$

By $L^2(\mu)$ we mean the completion of the inner product space obtained by identifying any two functions $f(x)$ and $g(x)$ such that

$$\int [g(t) - f(t)]^- d\mu(t) [g(t) - f(t)] = 0.$$

Then $L^2(\mu)$ inner product is written

$$\langle f, g\rangle_\mu = \int \bar{g}(t)\,d\mu(t) f(t).$$

Integrals of the form $\int \bar{g}(t)\,d\mu(t) f(t)$ are used also when either $f(x)$ or $g(x)$ is a Borel measurable, operator valued function. If for example $f(x)$ is vector valued and $g(x)$ is operator valued, the integral is taken to mean the unique vector c such that

$$\bar{a}c = \int [g(t)a]^- d\mu(t) f(t)$$

holds in the $L^2(\mu)$ sense for every vector a, if such a vector c exists. The result of integration is an operator if both $f(x)$ and $g(x)$ are operator valued.

In present applications we always have

$$(1) \qquad \int (1 + t^2)^{-1} d\tau(\mu(t)) < \infty.$$

We may then define a corresponding operator valued analytic function $\phi(z)$ for $y > 0$ and for $y < 0$ by requiring that its real part be a Poisson integral,

$$\operatorname{Re} \phi(x + iy) = \pi^{-1} y \int |t - z|^{-2} d\mu(t).$$

The arbitrary constant remaining in the definition of $\phi(z)$ in each half-plane is conveniently chosen consistently, so that $\bar{\phi}(z) = -\phi(\bar{z})$. The identity

$$(2) \qquad [\phi(z) + \bar{\phi}(w)]/[\pi i(\bar{w} - z)] = \int (t - z)^{-1}(t - \bar{w})^{-1} d\mu(t)$$

then results by a partial fraction decomposition when z and w are not real. Similarly, if ν is a non-negative, operator valued measure on the Borel sets of the real line, which satisfies (1), we may choose an operator valued function $\psi(z)$, defined and analytic for $y > 0$ and for $y < 0$, such that

$$(3) \qquad [\psi(z) + \bar{\psi}(w)]/[\pi i(\bar{w} - z)] = \int (t - z)^{-1}(t - \bar{w})^{-1} d\nu(t)$$

holds whenever z and w are not real. The generalized Hilbert transformation is a linear isometric transformation of $L^2(\mu)$ onto $L^2(\nu)$ which is defined if $\phi(z)$ and $\psi(z)$ can be chosen so as to satisfy the algebraic condition

$$\phi(z)I\psi(z)=I=\psi(z)I\phi(z) \tag{4}$$

whenever z is not real. If $f(x)$ is in $L^2(\mu)$, its generalized Hilbert transform is the unique element $g(x)$ of $L^2(\nu)$ such that the formula

$$\int (t-w)^{-1}g(t)\,d\nu(t)=\psi(w)I\int (t-w)^{-1}f(t)\,d\mu(t) \tag{5}$$

holds for all non-real values of w. The following theorem is a straightforward generalization of Theorem VIII of [3].

THEOREM I. *Let μ and ν be non-negative, operator valued measures on the Borel sets of the real line, which satisfy* (1) *and are related by* (4). *If $f(x)$ is any given element of $L^2(\mu)$, there exists a unique element $g(x)$ of $L^2(\nu)$ such that* (5) *holds for all non-real values of w. In this case,*

$$\int \bar{f}(t)\,d\mu(t)f(t)=\int \bar{g}(t)\,d\nu(t)g(t).$$

Every element $g(x)$ of $L^2(\nu)$ corresponds in this way to some element $f(x)$ of $L^2(\mu)$.

The situation of Theorem I may be considered a perturbation theory situation in two different, but equivalent, ways. First let us specify that we are concerned with the self-adjoint transformation, multiplication by x in $L^2(\mu)$, and with the self-adjoint transformation, multiplication by x in $L^2(\nu)$, which are defined by the formula $f(x)\to xf(x)$ in the two spaces. Because of the generalized Hilbert transformation, multiplication by x in $L^2(\nu)$ is equivalent to a self-adjoint transformation in $L^2(\mu)$, and we may compare the two self-adjoint transformations in this space. On the other hand, multiplication by x in $L^2(\mu)$ is equivalent to a self-adjoint transformation in $L^2(\nu)$, and we may compare the two self-adjoint transformations in this space. The two situations are of course equivalent via the generalized Hilbert transformation. Therefore the perturbation theory situation associated with Theorem I has an unambiguous meaning. As we will show it always satisfies the Kuroda perturbation hypothesis, and for this situation we will obtain the desired unitary equivalence by an explicit construction.

The conclusion of the theorem involves the decomposition of a self-adjoint transformation into absolutely continuous and singular parts, discussed by Kato [7]. To obtain a corresponding decomposition of the measure μ, let (e_n) be the choice of an orthonormal basis in $\mathcal{C}$. Then for every n, $\bar{e}_n\mu(E)e_n$

is a non-negative numerical measure when considered as a function of Borel sets E. Decompose each numerical measure into absolutely continuous and singular parts with respect to Lebesgue measure. Then we may construct two non-negative, operator valued measures μ_0 and μ_1, such that $\mu = \mu_0 + \mu_1$, and the numerical measures $\bar{e}_n\mu_0 e_n$ and $\bar{e}_n\mu_1 e_n$ are respectively absolutely continuous and singular for every n. In this case the trace measures $\tau(\mu_0)$ and $\tau(\mu_1)$ are absolutely continuous and singular respectively. This decomposition of μ does not depend on the choice of orthonormal set in $\mathfrak{K}$.

THEOREM II. *In the situation of Theorem* I, *multiplication by* x *in* $L^2(\mu_0)$ *is unitarily equivalent to the absolutely continuous part of multiplication by* x *in* $L^2(\mu)$.

The proof proceeds by an explicit construction. There is of course an analogous decomposition of ν and of $L^2(\nu)$. The Kuroda perturbation theorem now asserts the unitary equivalence of multiplication by x in $L^2(\mu_0)$ with multiplication by x in $L^2(\nu_0)$. This equivalence is obtained directly on multiplying by a suitable function, whose existence is now shown.

LEMMA 1. *Let* μ *be a non-negative, operator valued measure on the Borel sets of the real line which satisfies* (1) *and* (4), *and let* $\phi(z)$ *be a corresponding operator valued analytic function, defined so that* (2) *holds. Then,*

$$\phi_+(x) = \lim \phi(x + iy)$$

exists almost everywhere in the Schmidt norm as $y \searrow 0$.

In the situation of Theorem I, the conclusion of the Kuroda perturbation theorem can now be refined in the following way.

THEOREM III. *In the situation of Theorem* I,

$$f(x) \to I\bar{\phi}_+(x) f(x)$$

is a linear isometric transformation of $L^2(\mu_0)$ *onto* $L^2(\nu_0)$. *Multiplication by* x *in* $L^2(\mu_0)$ *is thereby unitarily equivalent to multiplication by* x *in* $L^2(\nu_0)$.

Functional notation is appropriate only in a dense subspace of $L^2(\mu_0)$, but the definition of the isometry is completed by continuity in the full space. The most general case of the Kuroda theorem will now be reduced to the situation just described.

As in Kato [7], we must first discard a portion of the given Hilbert space which is not relevant to the problem. If S and T are self-adjoint transformations in a Hilbert space $\mathfrak{H}$, let $\mathfrak{M}$ be the set of elements f of $\mathfrak{H}$

such that $(S-w)^{-1}f=(T-w)^{-1}f$ holds for all non-real values of w. Then, $\mathcal{M}$ is a closed vector subspace of $\mathfrak{H}$ which reduces both S and T. Since S and T coincide in $\mathcal{M}$, discussions of unitary equivalence may be restricted to the orthogonal complement of $\mathcal{M}$. In what follows, we may suppose, with no real loss of generality, that $\mathcal{M}$ is the zero subspace of $\mathfrak{H}$.

THEOREM IV. *Let S and T be self-adjoint transformations in a Hilbert space $\mathfrak{H}$ which satisfy the Kuroda hypothesis and there is no non-zero element f of $\mathfrak{H}$ such that $(S-w)^{-1}f=(T-w)^{-1}f$ holds for all non-real w. Let P_+ be the spectral projection corresponding to the half-plane $y>0$ for the normal operator*

$$1-(S-i)(S+i)^{-1}(T+i)(T-i)^{-1},$$

and let P_- be the specified projection corresponding to the half-plane $y<0$. Suppose that the dimension of (the range of) P_+ is no more than the dimension of $\mathcal{B}_+$, and that the dimension of the range of P_- is no more than the dimension of $\mathcal{B}_-$. Then there exist non-negative, operator valued measures μ and ν, which satisfy (1) *and are related by* (4), *and which have these properties*:

(1) *There exists a linear isometric transformation U of $\mathfrak{H}$ onto $L^2(\mu)$ which establishes the unitary equivalence of S with multiplication by x in $L^2(\mu)$.* (2) *There exists a linear isometric transformation V of $\mathfrak{H}$ onto $L^2(\nu)$ which establishes the unitary equivalence of T with multiplication by x in $L^2(\nu)$.* (3) *For every element f of $\mathfrak{H}$, the element Vf of $L^2(\nu)$ is the generalized Hilbert transform of Uf in $L^2(\mu)$.*

The proof proceeds by an explicit construction of U and V. Since $\mathcal{B}$ and I are arbitrary, Theorem IV applies, modulo the trivial space $\mathcal{M}$, to the most general situation in which the Kuroda hypothesis is valid. The four theorems taken together therefore give a sharpened form of the Kuroda theorem, which contains the Kato theorem, which in turn contains the Rosenblum theorem.

The proofs of our theorems require some lemmas involving Hilbert spaces of analytic functions. Similar spaces have previously been used in [5], where the notation, however, is different.

LEMMA 2. *Let μ be a non-negative, operator valued measure on the Borel sets of the real line which satisfies* (2). *Corresponding to every element $f(x)$ of $L^2(\mu)$, there exists a unique vector valued function $F(z)$, defined and analytic for $y>0$ and for $y<0$, such that*

$$\pi F(w)=\int(t-w)^{-1}d\mu(t)f(t) \tag{6}$$

holds for all non-real values of w. *A knowledge of* $F(z)$ *is sufficient to determine* $f(x)$ *uniquely. Let* $\mathcal{L}$ *be the Hilbert space of such vector valued analytic functions* $F(z)$, *with the norm defined so as to make the correspondence* $f(x) \rightarrow F(z)$ *isometric. If* $F(z)$ *is in* $\mathcal{L}$ *and if* w *is a non-real number, then* $[F(z) - F(w)]/(z - w)$ *again belongs to* $\mathcal{L}$, *and corresponds to* $f(x)/(x - w)$ *in* $L^2(\mu)$. *Let* $\phi(z)$ *be the choice of an operator valued analytic function which corresponds to* μ *according to* (2). *Then*

$$(7) \qquad [\phi(z) + \bar{\phi}(w)]c/[\pi i(\bar{w} - z)]$$

belongs to $\mathcal{L}$ *as a function of* z *for every vector* c *and for every non-real number* w, *and the formula*

$$(8) \qquad \bar{c}F(w) = \langle F(t), [\phi(t) + \bar{\phi}(w)]c/[\pi i(\bar{w} - t)]\rangle$$

holds for every $F(z)$ *in* $\mathcal{L}$. *If* w *is not real, the linear transformation of* $\mathcal{L}$ *into* $\mathcal{B}$ *defined by* $F(z) \rightarrow F(w)$ *is of Schmidt class. The space* $\mathcal{L}$ *contains no non-zero constant functions.*

Because of the identity (8), the Hilbert space $\mathcal{L}$ is uniquely determined by a knowledge of $\phi(z)$, and will be denoted in what follows as $\mathcal{L}(\phi)$. Similarly there is a Hilbert space $\mathcal{L}(\psi)$ which corresponds to $\psi(z)$ and $L^2(\nu)$. When $\mu = \mu_0 + \mu_1$ is decomposed into absolutely continuous and singular parts, we may define $\phi_0(z)$ and $\phi_1(z)$ by inserting subscripts in (2), and choosing the arbitrary constants consistently so that $\phi(z) = \phi_0(z) + \phi_1(z)$.

LEMMA 3. *Let* $\mu = \mu_0 + \mu_1$ *be a non-negative, operator valued measure which satisfies* (1). *Then,* $\mathcal{L}(\phi_0)$ *and* $\mathcal{L}(\phi_1)$ *are contained isometrically in* $\mathcal{L}(\phi)$ *as orthogonal subspaces which span* $\mathcal{L}(\phi)$.

The following lemma is inserted for completeness although it is not logically necessary to the discussion.

LEMMA 4. *The Kuroda hypothesis is satisfied under the hypotheses of Theorem* I.

For the proof of Theorem IV we need a characterization of the space of Lemma 2.

LEMMA 5. *Let* $\mathcal{L}$ *be a Hilbert space whose elements are vector valued functions, defined and analytic for* $y > 0$ *and for* $y < 0$. *Suppose that* $\mathcal{L}$ *contains no non-zero constant functions. Suppose that for every non-real number* w, *the linear transformation of* $\mathcal{L}$ *into* $\mathcal{B}$ *defined by* $F(z) \rightarrow F(w)$ *is of Schmidt class. Suppose also that* $[F(z) - F(w)]/(z - w)$ *belongs to* $\mathcal{L}$

whenever $F(z)$ *belongs to* $\mathcal{L}$, *for every non-real number* w, *and that the identity*

$$
\begin{aligned}
(9) \qquad 0 = {} & \langle F(t), [G(t) - G(\beta)]/(t-\beta)\rangle \\
& - \langle [F(t) - F(\alpha)]/(t-\alpha), G(t)\rangle \\
& + (\alpha - \bar{\beta})\langle [F(t) - F(\alpha)]/(t-\alpha), [G(t) - G(\beta)]/(t-\beta)\rangle
\end{aligned}
$$

holds whenever $F(z)$ *and* $G(z)$ *are in* $\mathcal{L}$, *and* α *and* β *are not real. Then* $\mathcal{L}$ *is equal isometrically to* $\mathcal{L}(\phi)$, *corresponding to some choice of non-negative, operator valued measure* μ *which satisfies* (1).

The Kuroda hypothesis and the dimension hypotheses of Theorem IV are used only for the following construction, which is extracted from the proof of the theorem.

LEMMA 6. *Under the hypotheses of Theorem* IV, *there exists a transformation* J *of* $\mathcal{H}$ *into* $\mathcal{C}$, *which is of Schmidt class and whose adjoint* J^* *takes* $\mathcal{C}$ *into* $\mathcal{H}$, *such that*

$$(10) \qquad 1 - (S-i)(S+i)^{-1}(T+i)(T-i)^{-1} = 2\pi J^* A J$$

holds for some unitary operator A, *which commutes with* I, *such that* $I(\bar{A} - A) \geqq 0$.

Proof of Lemma 1. Discussion of this limit requires the use of determinants for operators of the form $1 + A$, where A is of trace class. If A has finite dimensional range, this number is defined in terms of an orthonormal set $e_1, \cdots, e_r$ of vectors whose span contains the range of A and the orthogonal complement of its kernel. Then we may write $A = \sum A_{ij} e_i \bar{e}_j$, where $A_{ij} = \bar{e}_i A e_j$. Define $\det(1 + A)$ as the determinant of the $r \times r$-matrix $(\delta_{ij} + A_{ij})$. This definition does not depend on the choice of orthonormal set. Characteristic properties of the construction are:

$$
\begin{gathered}
\det(1 + \bar{A}) = [\det(1 + A)]^-, \\
[\det(1 + A)][\det(1 + B)] = \det(1 + A + B + AB), \\
\det(\exp A) = \exp(\tau(A)) \text{ if } A \geqq 0.
\end{gathered}
$$

There is now a unique way to extend the definition of $\det(1 + A)$ to operators A of trace class so as to maintain these properties, which imply continuity in the trace norm. We make particular use of the inequality

$$|\det(1 + A)|^2 = \det(1 + \bar{A} + A + \bar{A}A) \geqq 1,$$

which is valid if $A + \bar{A} \geqq 0$.

into $\mathcal{L}$. This adjoint may be written in the form $c \to K(w,z)c$, if $K(w,z)$ is a suitably chosen operator valued function of z, defined and analytic for $y > 0$ and for $y < 0$. By this construction, the identity

$$\bar{c}F(w) = \langle F(t), K(w,t)c\rangle$$

holds for every $F(z)$ in $\mathcal{L}$. Since $c \to K(w,z)c$ is a Schmidt transformation of $\mathcal{C}$ into $\mathcal{L}$,

$$\sum \| K(w,t)e_n \|^2 = \sum \bar{e}_n K(w,w) e_n < \infty$$

holds for any choice of orthonormal basis (e_n) in $\mathcal{C}$. In other words the non-negative operator $K(w,w)$ is of trace class. If α and β are non-real numbers, and if a and b are corresponding vectors, (9) yields the identity

$$\begin{aligned}\langle K(\alpha,t)a, [K(\beta,t)b - K(\beta,w)b]/(t-w)\rangle \\ = \langle [K(\alpha,t)a - K(\alpha,\bar{w})a]/(t-\bar{w}), K(\beta,t)b\rangle\end{aligned}$$

for all non-real values of w. Expanding according to the definition of $K(w,z)$ and simplifying, we obtain the operator identity

$$(\beta - \bar{\alpha})K(\alpha,\beta) = (\beta - \bar{w})K(w,\beta) - (\bar{\alpha} - \bar{w})K(\bar{w},\alpha)^{-}.$$

So we may choose an operator valued function $G(z)$, defined and analytic for $y > 0$ and for $y < 0$, such that

$$K(w,z) = [G(z) + \bar{G}(w)]/[\pi i(\bar{w} - z)]$$

holds for all non-real values of z and w. Since $K(w,w)$ is non-negative and of trace class, $G(w) + \bar{G}(w)$ is non-negative and of trace class in the half-plane $i(\bar{w} - w) > 0$. With the choice of an orthonormal basis (e_n) in $\mathcal{C}$, we may expand

$$\tfrac{1}{2}[G(w) + \bar{G}(w)] = \sum f_n(w) e_n \bar{e}_n,$$

where $f_n(x+iy) = \operatorname{Re} \bar{e}_n G(x+iy) e_n$ is a real valued function which is non-negative and harmonic for $y > 0$, and $f_n(x - iy) = -f_n(x+iy)$. By the Poisson representation of a function non-negative and harmonic in a half-plane, there exists a unique non-negative numerical measure μ_n on the Borel setas of the real line, and a unique non-negative number p_n, such that

$$f_n(x+iy) = \pi^{-1} y \int | t - z |^{-2} d\mu_n(t) + p_n y$$

holds for $y > 0$, and hence also by symmetry for $y < 0$. Let us construct the non-negative, operator valued measure $\mu = \sum \mu_n e_n \bar{e}_n$ and the non-negative operator $p = \sum p_n e_n \bar{e}_n$, so that we have the representation

$$K(w,z) = p + \int (t-z)^{-1}(t-\bar{w})^{-1} d\mu(t),$$

at least when $z = w$ is not real. The case in which z and w are distinct follows by a partial fraction decomposition. Since μ satisfies (1) by the trace finiteness of $K(w,w)$, the space $\mathcal{L}(\phi)$ is defined. By Lemma 2, this space contains no non-zero constants. Since

$$K(w,z) = p + [\phi(z) + \bar{\phi}(w)]/[\pi i(\bar{w} - z)],$$

the argument in the proof of Lemma 3 will show that $\mathcal{L}(\phi)$ is contained isometrically in $\mathcal{L}$, where its orthogonal complement is a space of constants. Since $\mathcal{L}$ contains no non-zero constants by hypothesis, $p = 0$ and $\mathcal{L}$ is equal isometrically to $\mathcal{L}(\phi)$.

Note. The proof of Lemma 5 contains an obvious error because the values of $G(z)$ do not commute. Corrections are easily made and are left to the reader to save a revision of the galley proof.

Proof of Lemma 6. The Kuroda hypothesis implies that the transformation appearing on the left of (10) is of trace class. Since this transformation is of the form $1 - W$, where W is a unitary transformation, it is diagonalizable. Let (f_n) be the choice of an orthonormal set in $\mathcal{H}$, consisting of eigenelements for $1 - W$, which spans the orthogonal complement of the kernel of this transformation. Then, $(1 - W)f_n = w_n f_n$ where $|1 - w_n| \leq 1$ for every n and $\sum |w_n| < \infty$. Our dimension hypotheses allow us to choose an orthogonal set (e_n) in $\mathcal{B}$ with e_n in $\mathcal{B}_+$ or $\mathcal{B}_-$ according as $i(\bar{w}_n - w_n)$ is positive or negative, the choice of $\mathcal{B}_+$ or $\mathcal{B}_-$ being immaterial if w_n is real. The transformation J is defined by $Jf_n = |2\pi w_n|^{\frac{1}{2}} e_n$, and is extended by linearity and continuity on the closed span of the set (f_n). It is defined to be zero on the orthogonal complement of this closed span. The adjoint transformation J^* then has the property that $J^* e_n = |2\pi w_n|^{\frac{1}{2}} f_n$. The operator A is defined by $Ae_n = |w_n|^{-1} w_n e_n$ for every n, and by linearity and continuity on the closed span of (e_n). It is taken to coincide with I on the orthogonal complement of this closed span. Formula (10) and the unitary property of A are obvious from this construction. The transformation J is of Schmidt class because $\sum |w_n| < \infty$. The operator A commutes with I because the constructed eigenvector (e_n) are eigenvectors of I. The operator $I(\bar{A} - A)$ is non-negative because the vectors (e_n) are eigenvectors of the operator for non-negative eigenvalues, and because the operator is zero on the orthogonal complement of the (e_n).

Proof of Theorem I. By Lemma 2, we must show that the transformation

$F(z) \to \psi(z)IF(z)$ takes $\mathcal{L}(\phi)$ isometrically onto $\mathcal{L}(\psi)$. To see this observe from (4) that the identity

$$(11) \qquad \psi(z) + \bar{\psi}(w) = -\psi(z)I[\phi(z) + \bar{\phi}(w)]I\bar{\psi}(w)$$

holds whenever z and w are not real. Therefore, if $F(z)$ is a finite linear combination of functions of the form (7), where c is in $\mathcal{C}$ and w is not real, then $\psi(z)F(z)$ belongs to $\mathcal{L}(\psi)$ and has the same norm as $F(z)$ in $\mathcal{L}(\phi)$. The necessary computation of inner products is based on the identity (8) in $\mathcal{L}(\phi)$ and its analogue in $\mathcal{L}(\psi)$. Such special functions are dense in $\mathcal{L}(\phi)$ because of (8). It follows by continuity that $F(z) \to \psi(z)IF(z)$ is a linear isometric transformation of $\mathcal{L}(\phi)$ into $\mathcal{L}(\psi)$. A similar argument from formula (16) will show that $G(z) \to -\phi(z)IG(z)$ is a linear isometric transformation of $\mathcal{L}(\psi)$ onto $\mathcal{L}(\phi)$. These two isometric transformations are inverses because of (4).

Proof of Theorem II. By Lemma 3, $\mathcal{L}(\phi)$ is the orthogonal direct sum of $\mathcal{L}(\phi_0)$ and $\mathcal{L}(\phi_1)$. The correspondence of Lemma 2 between $\mathcal{L}(\phi)$ and $L^2(\mu)$ induces an orthogonal decomposition of $L^2(\mu)$ into two subspaces which reduce multiplication by x in $L^2(\mu)$. The restriction of multiplication by x to each subspace is unitarily equivalent to multiplication by x in $L^2(\mu_i)$, $i = 0$ or 1. Since $\tau(\mu_0)$ is absolutely continuous and $\tau(\mu_1)$ is singular, with respect to Lebesgue measure, multiplication by x in $L^2(\mu_0)$ has an absolutely continuous spectral measure, in the sense of Rosenblum [12], and multiplication by x in $L^2(\mu_1)$ has a singular spectral measure. Therefore, our decomposition of operator measures corresponds to the spectral decomposition of Kato [7].

Proof of Theorem III. Since the trace measure $\tau(\mu)$ is a non-negative numerical measure, its derivative exists almost everywhere. It follows that

$$\mu'(x) = \lim \mu([x-h, x+h])/(2h)$$

exists almost everywhere in the trace norm as $h \searrow 0$. From the representation (2), one finds that

$$\mu'(x) = \lim \tfrac{1}{2}[\phi(x+iy) + \bar{\phi}(x+iy)]$$

holds almost everywhere in the trace norm as $y \searrow 0$. By Lemma 1,

$$\mu'(x) = \tfrac{1}{2}[\phi_+(x) + \bar{\phi}_+(x)]$$

holds for almost all real x. In a similar way,

$$\nu'(x) = \tfrac{1}{2}[\psi_+(x) + \bar{\psi}_+(x)]$$

holds almost everywhere. Passing to a limit in (11), we obtain

$$\nu'(x) = -\psi_+(x) I \mu'(x) I \bar{\psi}_+(x)$$

almost everywhere, and similarly also

$$\mu'(x) = -\phi_+(x) I \nu'(x) I \bar{\phi}_+(x),$$

where $I\bar{\phi}_+(x)$ and $-I\bar{\psi}_+(x)$ are inverse operators because of (4). The remainder of the proof is an exercise in integration which is left to the reader.

Proof of Theorem IV. Let J and A be defined as in Lemma 6. Corresponding to every element f of $\mathfrak{H}$, define a vector valued analytic function $F(z)$ by

$$F(w) = J(S-i)(S-w)^{-1} f \tag{12}$$

whenever w is not real. If such a function $F(z)$ vanishes identically, then $f = 0$. For by (10),

$$(S-i)(S-w)^{-1} f = (S-i)(S+i)^{-1}(T+i)(T-i)^{-1}(S-i)(S-w)^{-1} f$$

holds for all non-real values of w. So,

$$\begin{aligned}(S+i)(S-w)^{-1} f &= (T+i)(T-i)^{-1}(S-i)(S-w)^{-1} f \\ &= (S-i)(S-w)^{-1} f + 2i(T-i)^{-1}(S-i)(S-w)^{-1} f,\end{aligned}$$

and therefore,

$$\begin{aligned}(S-w)^{-1} f &= (T-i)^{-1}(S-i)(S-w)^{-1} f \\ &= (T-i)^{-1} f + (w-i)(T-i)^{-1}(S-w)^{-1} f.\end{aligned}$$

It follows that

$$(T-w)(T-i)^{-1}(S-w)^{-1} f = (T-i)^{-1} f$$

and that

$$(T-i)^{-1}(S-w)^{-1} f = (T-w)^{-1}(T-i)^{-1} f = (T-i)^{-1}(T-w)^{-1} f$$

and that $(S-w)^{-1} f = (T-w)^{-1} f$ holds for all non-real values of w. By hypothesis then, $f = 0$.

Make such functions $F(z)$ into a Hilbert space $\mathfrak{L}$ in such a way as to make the correspondence $f \to F(z)$ isometric. A partial fraction decomposition will show that $[F(z) - F(w)]/(z-w)$ belongs to $\mathfrak{L}$, if w is a non-real number, and correspond to $(S-w)^{-1} f$ in $\mathfrak{H}$. The identity (9) is satisfied in $\mathfrak{L}$ because of the self-adjointedness of S in $\mathfrak{H}$. Since S is a well-defined transformation, the transformation $(S-w)^{-1}$ has kernel zero when w is not real, and no non-zero constants can belong to $\mathcal{L}$. By this construction the linear transformation of $\mathcal{L}$ into $\mathcal{C}$ defined by $F(z) \to F(w)$ is of Schmidt class for every non-real choice of w. By Lemma 5, $\mathcal{L}$ is equal isometrically to $\mathcal{L}(\phi)$ corresponding to some choice of non-negative, operator valued measure μ which satisfies (2). A particular choice of the arbitrary constant in $\phi(z)$ is made below.

Since the transformation J of $\mathcal{H}$ into $\mathcal{C}$ corresponds to the transformation $F(z) \to F(i)$ of $\mathcal{L}(\phi)$ into $\mathcal{C}$, the adjoint transformation J^* of $\mathcal{C}$ into $\mathcal{H}$ corresponds to the adjoint transformation

$$c \to [\phi(z) + \bar{\phi}(i)]c/[\pi i(-i-z)]$$

of $\mathcal{C}$ into $\mathcal{L}(\phi)$. The transformation in $\mathcal{H}$ appearing on the left of (10) therefore corresponds to the transformation

$$F(z) \to [\phi(z) + \bar{\phi}(i)]2AF(i)/(1-iz)$$

in $\mathcal{L}(\phi)$, and the transformation

$$F(z) \to F(z) - [\phi(z) + \bar{\phi}(i)]2AF(i)/(1-iz)$$

is unitary in $\mathcal{L}(\phi)$. The adjoint is the transformation

$$F(z) \to F(z) - [\phi(z) + \bar{\phi}(i)]2\bar{A}F(i)/(1-iz).$$

Since the composition of the last two transformations is the identity transformation,

$$0 = [\phi(z) + \bar{\phi}(i)]\bar{A}[\phi(i) - \bar{A} + \bar{\phi}(i) - A]AF(i)$$

holds for every $F(z)$ in $\mathcal{L}(\phi)$. It follows that

$$\bar{c}[\phi(i) - \bar{A} + \phi(i) - A]c = 0$$

holds for every vector c in the closure of the range of the transformation $F(z) \to F(i)$. The same is true if c is orthogonal to this range since $\bar{c}[\phi(i) + \bar{\phi}(i)]c = 0$ because of (8) and $\bar{c}(A + \bar{A})c = 0$ if A is constructed as in the proof of Lemma 6. It follows that

$$\phi(i) - \bar{A} + \phi(i) - A = 0.$$

The arbitrary constant in the definition of $\phi(z)$ is now to be chosen so that $\phi(i) = \bar{A} = A^{-1}$.

With this normalization, the transformation in $\mathcal{H}$ appearing on the left of (10) corresponds to the transformation

$$F(z) \to [\phi(z) + \bar{\phi}(i)] 2\phi(i)^{-1} F(i)/(1 - iz)$$

in $\mathcal{L}(\phi)$. The transformation $(S - i)(S + i)^{-1}(T + i)(T - i)^{-1}$ in $\mathcal{H}$ therefore corresponds to the transformation

$$F(z) \to F(z) - [\phi(z) + \bar{\phi}(i)] 2\phi(i)^{-1} F(i)/(1 - iz)$$

in $\mathcal{L}(\phi)$. Since the transformation $(S + i)(S - i)^{-1}$ in $\mathcal{H}$ corresponds to

$$F(z) \to [(z + i)F(z) - 2iF(i)]/(z - i)$$

in $\mathcal{L}(\phi)$, the transformation $(T + i)(T - i)^{-1}$ in $\mathcal{H}$ corresponds to the transformation

$$F(z) \to [(z + i)F(z) - 2i\phi(z)\phi(i)^{-1}F(i)]/(z - i)$$

in $\mathcal{L}(\phi)$, and the transformation $(T - i)^{-1}$ in $\mathcal{H}$ corresponds to

$$F(z) \to [F(z) - \phi(z)\phi(i)^{-1}F(i)]/(z - i)$$

in $\mathcal{L}(\phi)$. Since $\phi(i)$ is an invertible operator and since $\phi(z)$ is continuous in the operator norm for $y > 0$, the operator $\phi(w)$ is invertible for w in some neighborhood of i. For such values of w, the transformation $(T - w)^{-1}$ in $\mathcal{H}$ must correspond to the transformation

$$F(z) \to [F(z) - \phi(z)\phi(w)^{-1}F(w)]/(z - w) \tag{13}$$

in $\mathcal{L}(\phi)$ because of the resolvent identity

$$(\alpha - \beta)(T - \alpha)^{-1}(T - \beta)^{-1} = (T - \alpha)^{-1} - (T - \beta)^{-1}.$$

By the self-adjointness of T and the symmetry $\bar{\phi}(w) = -\phi(\bar{w})$, formula (13) must also be valid for w in a neighborhood of $-i$.

The procedure used at the beginning of the proof is now used to construct a new Hilbert space $\mathcal{L}(\psi)$ of vector valued analytic functions, in a one-to-one isometric correspondence with $\mathcal{H}$. If g is in $\mathcal{H}$, the corresponding element $G(z)$ of $\mathcal{L}(\psi)$ is defined by

$$G(w) = I\phi(i)^{-1}J(T - i)(T - w)^{-1}g \tag{14}$$

whenever w is not real. The function $\psi(z)$ is of the form (3), where ν is a non-negative, operator valued measure which satisfies (1). A particular choice of the arbitrary constant in $\psi(z)$ is made below.

The defining isometric correspondences between $\mathcal{L}(\phi)$ and $\mathcal{H}$, and between $\mathcal{L}(\psi)$ and $\mathfrak{H}$, induce an isometric correspondence between $\mathcal{L}(\phi)$ and $\mathcal{L}(\psi)$. Let $F(z)$ be in $\mathcal{L}(\phi)$ and let $G(z)$ be the corresponding element of $\mathcal{L}(\psi)$. When w is in a neighborhood of i or $-i$, the action of $(T-i)(T-w)^{-1}$ corresponds to

$$F(z) \rightarrow [(z-i)F(z)-(w-i)\phi(z)\phi(w)^{-1}F(w)]/(z-w)$$

in $\mathcal{L}(\phi)$, and the action of $J(T-i)(T-w)^{-1}$ corresponds to

$$F(z) \rightarrow \phi(i)\phi(w)^{-1}F(w).$$

The definition (14) therefore implies that $G(w)=I\phi(w)^{-1}F(w)$ hold when w is a neighborhood of i or $-i$. When this relation is written in the form

$$F(z)=-\phi(z)IG(z), \tag{15}$$

it is valid for all non-real values of z by analytic continuation.

Now use formula (8) in $\mathcal{L}(\phi)$ and its analogue in $\mathcal{L}(\psi)$. The isometric nature of the correspondence (15) between $\mathcal{L}(\phi)$ and $\mathcal{L}(\psi)$ implies that

$$\phi(z)+\bar{\phi}(w)=-\phi(z)I[\psi(z)+\bar{\psi}(w)]I\bar{\phi}(w) \tag{16}$$

holds whenever z and w are not real. Since $\phi(i)$ is a unitary operator, this identity implies that

$$[\phi(i)+I\bar{\psi}(i)I]+[\bar{\phi}(i)+I\psi(i)I]=0.$$

We may therefore choose the arbitrary constant in $\psi(z)$ so that $\phi(i)+I\bar{\psi}(i)I=0$. Formula 16) now implies that $\phi(z)I\psi(z)=I$. Since $\phi(z)$ is known to have invertible values in a neighborhood of i and in a neighborhood of $-i$, we must also have $\psi(z)I\phi(z)=I$ for these values of z, and hence by analytic continuation whenever z is not real. Formula (4) has now been obtained.

By Lemma 2 and the proof of Theorem I, we now have a commutative diagram of isometric correspondences between the spaces $\mathfrak{H}$, $\mathcal{L}(\phi)$, $\mathcal{L}(\psi)$, $L^2(\mu)$, and $L^2(\nu)$, where the linear isometric transformation of $L^2(\mu)$ onto $L^2(\nu)$ is the generalized Hilbert transformation. The linear isometric transformation of $L^2(\mu)$ onto $\mathcal{L}(\phi)$ is given by (6). The linear isometric transformation U of $\mathfrak{H}$ onto $L^2(\mu)$ is the inverse of this transformation composed with the transformation of $\mathfrak{H}$ into $\mathcal{L}(\phi)$ given by (12). The linear isometric transformation V of $\mathfrak{H}$ into $L^2(\nu)$ is defined analogously. That Vf is the generalized Hilbert transform of Uf for every f in $\mathfrak{H}$ is clear from the proof of Theorem I.

NEW YORK UNIVERSITY.

REFERENCES.

[1] L. de Branges, "Some Hilbert spaces of entire functions," *Transactions of the American Mathematical Society*, vol. 96 (1960), pp. 259-295.

[2] ———, "Some Hilbert spaces of entire functions II," *Transactions of the American Mathematical Society*, vol. 99 (1961), pp. 118-152.

[3] ———, "Some Hilbert spaces of entire functions III," *Transactions of the American Mathematical Society*, vol. 100 (1961), pp. 73-115.

[4] ———, "Some Hilbert spaces of entire functions IV," *Transactions of the American Mathematical Society*, to appear.

[5] ———, "Some Hilbert spaces of analytic functions," *Transactions of the American Mathematical Society*, to appear.

[6] K. O. Friedrichs, "On the perturbation of the continuous spectra," *Communications on Pure and Applied Mathematics*, vol. 1 (1948), pp. 361-406.

[7] T. Kato, "On finite-dimensional perturbations of self-adjoint operators," *Journal of the Mathematical Society of Japan*, vol. 9 (1957), pp. 239-249.

[8] ———, "Perturbations of continuous spectra by trace class operators," *Proceedings of the Japan Academy*, vol. 33 (1957), pp. 260-264.

[9] S. T. Kuroda, "Perturbations of continuous spectra by unbounded operators I," *Journal of the Mathematical Society of Japan*, vol. 11 (1959), pp. 247-262.

[10] ———, "Perturbations of continuous spectra by unbounded operators II," *Journal of the Mathematical Society of Japan*, vol. 12 (1960), pp. 243-257.

[11] ———, "On the existence and the unitary property of the scattering operator," *Il Nuovo Cimento*, vol. 12 (1959), pp. 431-454.

[12] M. Rosenblum, "Perturbation of the continuous spectrum and unitary equivalence," *Pacific Journal of Mathematics*, vol. 7 (1957), pp. 997-1010.

[13] ———, "The absolute continuity of Toeplitz's matrices," *Pacific Journal of Mathematics*, vol. 10 (1960), pp. 987-996.

[14] R. Schatten, *Norm ideals of completely continuous operators*, Springer-Verlag, Berlin, 1960.

FUNDAMENTAL SOLUTIONS OF LINEAR PARTIAL DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS DEPENDING ON PARAMETERS.*

By François Treves.[1]

We deal with a linear partial differential operator

$$P(\lambda, D_x) = \sum_{|p| \leq m(\lambda)} A_p(\lambda) D_x{}^p$$

whose coefficients $A_p(\lambda)$, constant with respect to the variable $x = (x_1, \cdots, x_n) \in R^n$, are complex functions of the point λ of a C^∞ manifold Λ.[2] As usually $p = (p_1, \cdots, p_n)$ is a multi-index, the p_j being non negative integers, $|p| = p_1 + \cdots + p_n$ and $D^p = \left(\frac{1}{i}\frac{\partial}{\partial x_1}\right)^{p_1} \cdots \left(\frac{1}{i}\frac{\partial}{\partial x_n}\right)^{p_n} (i = \sqrt{-1})$. The *order* $m(\lambda)$ of $P(\lambda, D_x)$ will be allowed to vary with λ. However, we shall assume that it is locally bounded in Λ, i.e. that $m(\lambda)$ is bounded on every compact subset of Λ.

We say that $P(\lambda, D_x)$ does not vanish in Λ if its coefficients $A_p(\lambda)$ do not vanish simultaneously for any $\lambda \in \Lambda$.

The main result of this article is then the following:

Theorem 1. *Suppose that $P(\lambda, D)$ does not vanish in Λ and that its coefficients are C^∞ functions of λ. Then there is a C^∞ function $E(\lambda)$ of λ with values in the space of distributions $\mathcal{D}'_x$ on R^n which, for each $\lambda \in \Lambda$, is a fundamental solution of $P(\lambda, D)$.*

The conclusion of Theorem 1 can be strengthened by strengthening the conditions upon $P(\lambda, D_x)$. Indeed:

Theorem 2. *Same hypotheses as in Theorem 1. Assume, moreover, that for any two values λ', λ'' of λ, the differential operators $P(\lambda', D_x)$, $P(\lambda'', D_x)$ are equally strong. Then there is a C^∞ function of $\lambda \in \Lambda$, $E(\lambda)$, with values in $(\mathcal{F}^{-1}L^\infty)_{loc}$, which is, for each λ, a fundamental solution of $P(\lambda, D_x)$.*

The space $(\mathcal{F}^{-1}L^\infty)_{loc}$, often denoted by $H_{loc}{}^{\infty,0}$ or $H_{\infty,0}{}^{loc}$ is the space of distributions T on R^n such that, for every $C_0{}^\infty$ function $\phi(x)$, the Fourier

---

* Received February 19, 1962.

[1] Sloan fellow.

[2] By a manifold, we mean here a *paracompact* and *Hausdorff* manifold.

transform $\mathcal{F}(\phi T)$ of ϕT belongs to L^∞. It is provided with the topology defined by the semi-norms

$$T \to \| \mathcal{F}(\phi T) \|_{L^\infty}, \phi \in C_0^\infty.$$

The importance of $(\mathcal{F}^{-1}L^\infty)_{loc}$ comes from the fact that its elements are exactly those distributions T such that

$$T * L^2{}_c \subset L^2{}_{loc}$$

where $*$ is the convolution product, $L^2{}_c$ the space of L^2-functions with compact support, $L^2{}_{loc}$ the space of locally L^2-functions. Furthermore, if one provides the space $\mathcal{L}(L^2{}_c; L^2{}_{loc})$ of continuous linear mappings $L^2{}_c \to L^2{}_{loc}$ with the topology of uniform convergence on the bounded subsets of $L^2{}_c$, the map $T \to T*$ is an isomorphism of $(\mathcal{F}^{-1}L^\infty)_{loc}$ into $\mathcal{L}(L^2{}_c : L^2{}_{loc})$.

Let us recall also the meaning of "equally strong." If $P(\xi)$ is a polynomial with complex coefficients in n variables $\xi_1, \cdots, \xi_n$, we associate with it a differential operator with constant coefficients (or differential polynomial) $P(D)$, by substituting $\frac{1}{i}\frac{\partial}{\partial x_\alpha}$, $\alpha = 1, \cdots, n$, for ξ_α. We also associate with $P(\xi)$ the quantity

$$\tilde{P}(\xi) = (\sum_p | P^{(p)}(\xi)|^2)^{\frac{1}{2}}, \xi \in R^n,$$

where the summation runs over all systems $p = (p_1, \cdots, p_n)$ of n non-negative integers $p_1, \cdots, p_n$, and where

$$P^{(p)}(\xi) = \left(\frac{\partial}{\partial \xi_1}\right)^{p_1} \cdots \left(\frac{\partial}{\partial \xi_n}\right)^{p_n} P(\xi).$$

With those notations, $P(D)$ is said to be *stronger* than $Q(D)$ if for some $c > 0$ we have, for all $\xi \in R^n$,

$$c\tilde{Q}(\xi) < \tilde{P}(\xi).$$

And $P(D)$, $Q(D)$ are said to be *equally strong* if each one is stronger than the other.[3]

Our way of proving Theorem 2 will also yield the following result.

THEOREM 3. *Same hypotheses as in Theorem 2. Assume, moreover, that* Λ *is provided with a structure of complex analytic manifold and that the coefficients of* $P(\lambda, D_x)$ *are holomorphic functions of* $\lambda \in \Lambda$.

Then, every point $\lambda^0 \in \Lambda$ *has an open neighborhood* $N(\lambda^0)$ *such that there is a holomorphic function of* $\lambda \in N(\lambda^0)$, $E(\lambda)$, *with values in* $(\mathcal{F}^{-1}L^\infty)_{loc}$, *which is, for each* $\lambda \in N(\lambda^0)$, *a fundamental solution of* $P(\lambda, D_x)$.

---

[3] On these notions, see Hörmander [1].

We are going to prove Theorems 1, 2, 3 by constructing $E(\lambda)$. By this construction we shall obtain a fundamental solution $E(\lambda)$ endowed with rather strong properties.

I. Proof of Theorem 1.

The result to be proved is of purely local nature. More precisely, assume that we have a locally finite open covering $\{\Lambda^i\}$ of the manifold Λ and that for each i, we have found a fundamental solution $E^i(\lambda)$ of $P(\lambda, D)$ for $\lambda \in \Lambda^i$ such that $\lambda \to E^i(\lambda)$ is C^∞ with values in $\mathcal{D}'_x$. Let $\{a^i(\lambda)\}$ be a partition of unity in $C_0^\infty(\Lambda)$ subordinated to the covering $\{\Lambda^i\}$. Then

$$E(\lambda) = \sum_i a^i(\lambda) E^i(\lambda)$$

has the desired properties. It will, moreover, be fairly evident that the additional properties which will be proved about the $E^i(\lambda)$'s, will also be valid for $E(\lambda)$.

Such a localization with respect to λ allows us to assume that the order $m(\lambda)$ of $P(\lambda, D_x)$ is bounded by an integer m; of course, we take $m \geqq 1$. Otherwise $m(\lambda)$ can vary from 0 to m when λ varies.

We can take now, as Λ, an open ball centered at the origin in the euclidean space R^d (if $d = \dim \Lambda$) and assume that $P(\lambda, D_x)$ has all the required properties for λ in some strictly larger ball Λ': the $A_p(\lambda)$ are C^∞ functions of $\lambda \in \Lambda'$; they cannot vanish simultaneously in Λ'.

I. 1. Two lemmas about polynomials.

LEMMA 1. *Let $\epsilon > 0$ be arbitrary. There is a finite set $\mathfrak{S}_\epsilon$ of vectors $\theta \in R^n$, with $|\theta| < \epsilon$, and a constant $c > 0$ such that, for all polynomials P in n variables of degree $\leqq m$, and all $\xi \in R^n$,*

$$c\tilde{P}(\xi) \leqq \sup_{\theta \in \mathfrak{S}_\epsilon} |P(\xi + \theta)|.$$

Let $N(m, n)$ be the set of multi-indices $p = (p_1, \cdots, p_n)$, p_j: integer $\geqq 0$, such that $|p| = p_1 + \cdots + p_n \leqq m$. We order $N(m, n)$ in an arbitrary way. For each vector $\theta = (\theta_1, \cdots, \theta_n) \in R^n$ and $q \in N(m, n)$, set $\theta^q = (\theta_1)^{q_1} \cdots (\theta_n)^{q_n}$. Consider the mapping

$$(\theta_p) \to (\theta_p{}^q)$$

where $\theta_p = (\theta_{p,1}, \cdots, \theta_{p,n}) \in R^n$. That mapping is defined in $(R^n)^{N(m,n)}$ and valued in the linear space of $r(m, n) \times r(m, n)$ matrices ($r(m, n)$: number of element of $N(m, n)$). The set $\mathfrak{V}$ of elements $(\theta_p) \in (R^n)^{N(m,n)}$ such that

$\det(\theta_p{}^q) = 0$, i. e. such that the matrix $(\theta_p{}^q)$ is not invertible, is an algebraic set. It is easy to see that $\mathcal{V}$ cannot be the whole space. It follows from this that $\mathcal{V}$ contains points (θ_p) arbitrarily close to the origin, hence such that $|\theta_p| < \epsilon$ for all $p \in N(m,n)$. Let (θ_p) be such and call $\mathfrak{S}_\epsilon$ the set of the "components" θ_p. In view of Taylor's formula

$$P(\xi+\theta_p) = \sum_{q\in N(m,n)} \theta_p{}^q \frac{P^{(q)}(\xi)}{q!},\ p \in N(m,n).$$

Since the matrix $(\theta_p{}^q)$ is invertible, there is $M < +\infty$ such that

$$\sum_q 1/q!\,|P^{(q)}(\xi)| \leqq M \sum_q |P(\xi+\theta_p)|.$$

Since M depends only on $\mathfrak{S}_\epsilon$ the lemma is proved.

The following lemma will play an essential role in the sequel.

Lemma 2. *Let $\epsilon > 0$ be arbitrary. There is a finite set of vectors $\mathfrak{S}'_\epsilon \subset R^n$, such that $|\theta| < \epsilon$ if $\theta \in \mathfrak{S}'_\epsilon$, and a constant $c > 0$ such that for all polynomials P in n variables of degree $\leqq m$, and for all $\xi \in R^n$,*

$$c'P(\xi) \leqq \sup_{\theta\in\mathfrak{S}_\epsilon} \inf_{|z|=1} |P(\xi+z\theta)|.$$

Here z is a complex variable.

Let us fix ξ, $\theta \in R^n$ and set $g(z) = P(\xi+z\theta)$, $m' = \deg g(\leqq m)$. We have:

$$|g(z)| = |A| \prod_{j=1}^{m'} |z-z_j| \geqq |g(1)| \prod_{j=1}^{m'} \frac{||z|-|z_j||}{1+|z_j|}.$$

For some $k = 0, 1, \cdots, m$ at least, we have

$$|k/m - |z_j|| \geqq \frac{1}{2m+1} \text{ for all } j = 1, \cdots, m' \leqq m.$$

Take $|z| = k/m$ for such a k. If $|z_j| \geqq 2$, $\dfrac{||z|-|z_j||}{1+|z_j|} \geqq \frac{1}{3}$. If $|z_j| < 2$, $\dfrac{||z|-|z_j||}{1+|z_j|} \geqq \frac{1}{3}\dfrac{1}{2m+1}$. Therefore, we have in all cases

$$|P(\xi+z\theta)| \geqq \left[\frac{1}{3(2m+1)}\right]^m |P(\xi+\theta)|.$$

Combining with Lemma 1 we see that the set

$$\mathfrak{S}'_\epsilon = \{(k/m)\theta \mid k = 0, 1, \cdots, m; \theta \in \mathfrak{S}_\epsilon\}$$

satisfies the conditions of Lemma 2. We may take

$$c' = c\left[\frac{1}{3(2m+1)}\right]^m$$

These reasonings and results are due to L. Hörmander. From now on we keep $\epsilon > 0$ fixed (but arbitrary) and we choose $\mathfrak{S}'_\epsilon$, which we shall call S, once forever.

I. 2. When the parameters vary. We go back to the differential polynomials $P(\lambda, D_x)$. We introduce now a continuous function $R(\xi) \geqq 0$ on R^n with the following properties:

i) For $A > 0$ sufficiently large,

$$R(\xi) \leqq A(1 + |\xi|)^m, \ \xi \in R^n.$$

ii) There is a constant $C < +\infty$ such that, for all ξ, $\xi^0 \in R^n$, with $|\xi - \xi^0| \leqq 1$,

$$R(\xi) \leqq CR(\xi^0).$$

iii) For every multi-index $r = (r_1, \cdots, r_n)$ there is a constant $B(|r|) < +\infty$ such that for multi-indices $p = (p_1, \cdots, p_n)$, all $\lambda \in \Lambda$ and all $\xi \in R^n$,

$$|D_\lambda{}^r P^{(p)}(\lambda, \xi)| \leqq B(|r|) R(\xi).^4 \tag{I.0}$$

In view of (I.0) we see that there is a constant $B < +\infty$ such that for all λ, $\lambda^0 \in \Lambda$ and $\xi \in \boldsymbol{R}^n$

$$|P^{(p)}(\lambda, \xi) - P^{(p)}(\lambda^0, \xi)| \leqq B |\lambda - \lambda^0| R(\xi). \tag{I.1}$$

In turn, (I.1) implies

$$|P(\lambda, \xi + z\theta) - P(\lambda^0, \xi + z\theta)| \leqq Be^{|z\theta|} |\lambda - \lambda^0| R(\xi) \tag{I.2}$$

and also

$$|P(\lambda, \xi) - P(\lambda^0, \xi)| \leqq B |\lambda - \lambda^0| R(\xi). \tag{I.3}$$

Lemma 3. *There are constants $\eta > 0$, $a > 0$, $b > 0$ such that, whatever be $\lambda^0 \in \Lambda$, $\xi^0 \in R^n$, one can find $\theta \in S$ with the property that*

$$|\xi - \xi^0| \leqq \eta, \ |\lambda - \lambda^0| \leqq a\tilde{P}(\lambda^0, \xi^0)/R(\xi^0)$$

implies

$$b\tilde{P}(\lambda, \xi) \leqq \inf_{|z|=1} |P(\lambda, \xi + z\theta)|.$$

We recall that S is the set of vectors $\mathfrak{S}'_\epsilon$ chosen at the end of Section I. 2.

According to Lemma 2, there is $\theta \in S$ such that

$$c'\tilde{P}(\lambda^0, \xi^0) \leqq \inf_{|z|=1} |P(\lambda^0, \xi^0 + z\theta)|. \tag{I.4}$$

---

[4] That such functions $R(\xi)$ exist is obvious: just take $R(\xi) = (1 + |\xi|)^m$.

Let us first ask that η be < 1. Then, in view of Taylor's formula, for $|\xi - \xi^0| < \eta$,

$$\text{(I.5)} \qquad |\tilde{P}(\lambda, \xi) - \tilde{P}(\lambda, \xi^0)| \leqq A\eta \tilde{P}(\lambda, \xi),$$

$$\text{(I.6)} \qquad |P(\lambda, \xi + z\theta) - P(\lambda, \xi^0 + z\theta)| \leqq A e^{\epsilon} \eta \tilde{P}(\lambda, \xi),$$

taking into account the fact that $|z\theta| < \epsilon$. Here $A < +\infty$ depends only on m. Combining (I.4) with (I.5) and (I.6) where $\lambda = \lambda^0$, we obtain

$$\text{(I.7)} \qquad \inf_{|z|=1} |P(\lambda^0, \xi + z\theta)| \geqq [c' - (c' + e^{\epsilon}) A\eta] \tilde{P}(\lambda^0, \xi).$$

We may choose $\eta \leqq \dfrac{1}{2A} \dfrac{c'}{c' + e^{\epsilon}}$. We combine then (I.7) with the inequalities (I.2) and (I.3); it yields

$$\inf_{|z|=1} |P(\lambda, \xi + z\theta)| \geqq (c'/2)\tilde{P}(\lambda, \xi) - ((c'/2) + e^{\epsilon}) |\lambda - \lambda^0| BR(\xi);$$

but $R(\xi) \leqq CR(\xi^0)$ since $|\xi - \xi^0| < 1$. Now take $|\lambda - \lambda^0| < a\tilde{P}(\lambda^0, \xi^0)/R(\xi^0)$. This gives

$$\text{(I.8)} \qquad \inf_{|z|=1} |P(\lambda, \xi + z\theta)| \geqq (c'/2)\tilde{P}(\lambda, \xi) - a((c'/2) + e^{\epsilon}) CB\tilde{P}(\lambda^0, \xi^0).$$

On the other hand, by applying (I.3) and (I.5),

$$\begin{aligned} \tilde{P}(\lambda^0, \xi^0) &\leqq \tilde{P}(\lambda, \xi) + CB|\lambda - \lambda^0| R(\xi^0) + A\tilde{P}(\lambda, \xi) \\ &\leqq aCB\tilde{P}(\lambda^0, \xi^0) + (1 + A)\tilde{P}(\lambda, \xi). \end{aligned}$$

We ask then that $aCB < 1$. This leads to

$$\tilde{P}(\lambda^0, \xi^0) \leqq \frac{1 + A}{1 - aCB} \tilde{P}(\lambda, \xi).$$

Replacing in (I.8)

$$\inf_{|z|=1} |P(\lambda, \xi + z\theta)| \geqq [(c'/2) - a((c'/2) + e^{\epsilon}) \frac{CB(1 + A)}{1 - aCB}] \tilde{P}(\lambda, \xi)$$

from where the possibility of finding $a > 0$, $b > 0$ as in Lemma 3 follows immediately.

Remark I.1. It is easily checked, on the proof of Lemma 3, that one may keep the same numbers η, a, b if one replaces $R(\xi)$ by any other function (continuous and non-negative) enjoying the properties i), ii), iii) with the same constants A, C, $B(|r|)$ as $R(\xi)$.

LEMMA 4. *Set* $h(\xi) = \inf_{\lambda\in\Lambda} \tilde{P}(\lambda,\xi)$. *There is a constant* $c > 0$ *such that*

$$h(\xi) \geqq c(1+|\xi|)^{-m}, \xi \in R^n.$$

From Taylor's formula,

$$\tilde{P}(\lambda,0) \leqq C(1+|\xi|)^m \tilde{P}(\lambda,\xi)$$

where C depends only on the upper bound m for the degree of $P(\lambda,\xi)$ for $\lambda\in\Lambda$ and on the dimension n of the ξ-space. Since $\tilde{P}(\lambda,0)$ is a continuous > 0 function of $\lambda\in\bar{\Lambda}$, the lemma follows.

I. 3. Partitions of unity. We shall denote by $k = (k_1,\cdots,k_d)$ a point of R^d with integral $\geqq 0$ or < 0 coordinates. We denote by $B(k)$ the open ball of radius $\sqrt{d}$ and center k. When k varies, the $B(k)$ form an open covering of R^d. There is an integer ν, depending only on d, such that any point of R^d is contained in at most ν balls $B(k)$. Let τ be > 0 and let us look at the automorphism $y \to y/\tau$ of R^d. The ball $B(k)$ is mapped onto the ball $B_\tau(k)$ of center k/τ and radius $\sqrt{d}/\tau$. The balls $B_\tau(k)$ form an open covering of R^d and at most ν of them meet.

Let $\alpha(t)$ be a C^∞ function of $t \geqq 0$,

$$\begin{aligned} \alpha(t) &= 0 && \text{for } t \geqq 1; \\ \alpha(t) &= 1 && \text{for } 0 \leqq t \leqq \tfrac{1}{2}; \\ 0 < \alpha(t) &\leqq 1 && \text{for } \tfrac{1}{2} \leqq t < 1. \end{aligned}$$

For $k = (k_1,\cdots,k_d)$ with integral coordinates we set

$$\alpha^k(y;\tau) = \alpha\left(\frac{|y-k/\tau|}{\sqrt{d}/\tau}\right),\ g(y;\tau) = \sum_k \alpha^k(y;\tau).$$

The support of $\alpha^k(y;\tau)$ is the closed ball $B_\tau(k)$. For every integer $\mu \geqq 0$, there is a constant $M(\mu) < +\infty$, depending only on α and μ (hence not on y, τ, k) such that for $|p| \leqq \mu$,

$$|D_y{}^p \alpha^k(y;\tau)| \leqq M(\mu)\tau^\mu \quad (\tau \geqq 1).$$

Since $\alpha^k(y;\tau)$ vanishes of infinite order on the boundary of $B_\tau(k)$ and since at most ν balls $B_\tau(k)$ intersect, we have

$$|D_y{}^p g(y;\tau)| \leqq \nu M(|p|)\tau^{|p|}.$$

We set then

$$\tilde{\alpha}^k(y;\tau) = \alpha^k(y;\tau)/g(y;\tau).$$

We have

$$\sum_k | D_y^p \tilde{\alpha}^k(y;\tau) | \leqq M'(| p |)\tau^{|p|}$$

where $M'(| p |)$ depends only on α and $| p |$. The $\tilde{\alpha}^k(y;\tau)$ form a partition of unity in $C_0^\infty(R^d)$.

Let $\eta > 0$ be as in Lemma 3. We choose a sequence $\{\xi^j\}$ of points in R^n such that the open balls $| \xi - \xi^j | < \eta$ form a locally finite (open) covering of R^n. Let $\{\beta^j\}$ be a partition of unity in $C_0^\infty(R^n)$ subordinated to this covering.

Fix $j = 0, 1, 2, \cdots$ and note that, according to our assumptions, $\lambda \to \tilde{P}(\lambda, \xi^j)$ is a continuous > 0 function on $\bar{\Lambda} \subset R^d$ and therefore reaches its minimum $h(\xi^j) > 0$ in Λ. Let us define τ^j by

$$\sqrt{d}/\tau^j = \inf[a \frac{h(\xi^j)}{R(\xi^j)}, \tfrac{1}{2} \text{ radius of } \Lambda].$$

Observe that there is a vector $\theta^{jk} \in S$ such that

$$| \xi - \xi^j | < \eta, | \lambda - k/\tau^j | \leqq \sqrt{d}/\tau^j, \lambda, k/\tau^j \in \Lambda,$$

implies

$$b\tilde{P}(\lambda, \xi) \leqq \inf_{|z|=1} | P(\lambda, \xi + z\theta^{jk}) |.$$

We set then

$$\gamma^{jk}(\xi, \lambda) = \beta^j(\xi)\tilde{\alpha}^k(\xi;\tau^j).$$

We have

$$\sum_{j,k} \gamma^{jk}(\xi, \lambda) = \sum_j \beta^j(\xi) \sum_k \tilde{\alpha}^k(\lambda;\tau^j) = \sum_j \beta^j(\xi) = 1$$

for all $\xi \in R^n$, $\lambda \in R^d$. Thus $\{\gamma^{jk}\}$ is a partition of unity in $C_0^\infty(R^n \times R^d)$. Note that because of the term $\frac{1}{2}$ radius of Λ in the definition of τ^j, the open ball $\frac{1}{2}\Lambda \subset R^d$ does not intersect any ball $| \lambda - k/\tau^j | \leqq \sqrt{d}/\tau^j$ whose center k/τ^j lies outside Λ. Therefore, for $\lambda \in \frac{1}{2}\Lambda$, $(\xi, \lambda) \in$ support of γ^{jk} implies

$$b\tilde{P}(\lambda, \xi) \leqq \inf_{|z|=1} | P(\lambda, \xi + z\theta^{jk}) |. \tag{I.9}$$

I. 4. Construction of $E(\lambda)$ in a neighborhood of a point λ^0. In fact we shall assume, as before, that Λ is an open ball in R^d centered at the origin. We shall take $\lambda^0 = 0$ and the neighborhood of λ^0 will be $\frac{1}{2}\Lambda$.

For $j = 0, 1, 2, \cdots$, and $k = (k_1, \cdots, k_d)$ with integral coordinates, we set

$$\langle E^{jk}(\lambda), \check{\phi} \rangle = \int \gamma^{jk}(\xi, \lambda) \int\int_{|z|=1} \frac{\hat{\phi}(\xi + z\theta^{jk})}{P(\lambda, \xi + z\theta^{jk}) \cdot 2i\pi z} \frac{dz}{} d\xi.^{5}$$

---

[5] The idea of using such a formula to construct a fundamental solution, when there are no parameters, is due to L. Hörmander.

Here $\lambda \in \Lambda/2$ and $\phi \in C_0^\infty(R^n)$. We have set

$$\hat{\phi}(\zeta) = \int \phi(x) e^{-i\langle x,\zeta\rangle} dx, \zeta \in C^n,$$

and $\check{\phi}(x) = \phi(-x)$. The integral with respect to z is taken over the unit circle $|z| = 1$ of the complex plane. It is easily seen that for $\lambda \in \Lambda/2$, $E^{jk}(\lambda)$ is a distribution on R_x^n. But we want to show much more. Among other things, we are going to show that $(1/(2\pi)^n) \sum_{j,k} E^{jk}(\lambda)$ converges in $\mathcal{D}'_x$ (space of distributions on R^n) to a C^∞ function of $\lambda \in \frac{1}{2}\Lambda$ with values in $\mathcal{D}'_x$.

Let $r = (r_1, \cdots, r_d)$, $|r| = \mu$. We call $F^{jk}(\lambda)$ the function of $\lambda \in \frac{1}{2}\Lambda$, $\langle E^{jk}(\lambda), \check{\phi}\rangle$. We have with obvious notations

$$D_\lambda{}^r F^{jk}(\lambda) = \sum_{s_1=0}^{r_1} \cdots \sum_{s_d=0}^{r_d} \binom{r}{s} \int D_\lambda{}^{r-s} \gamma^{jk}(\xi,\lambda)$$
$$\times \int_{|z|=1} D_\lambda{}^s \left[\frac{1}{P(\lambda, \xi + z\theta^{jk})}\right] \check{\phi}(\xi + z\theta^{jk}) \frac{dz}{2i\pi z}\, d\xi.$$

First of all,

$$\sum_k | D_\lambda{}^{r-s} \gamma^{jk}(\xi,\lambda)|$$
$$= \beta^j(\xi) \sum_k | D_\lambda{}^{r-s} \tilde{\alpha}^k(\lambda; \tau^j)| \leqq \beta^j(\xi) M'(|r-s|)(\tau^j)^{|r-s|}.$$

On the other hand, in view of (I. 0) there is a constant $M''(|s|)$ depending only on $|s|$, such that, for (ξ, λ) in the support of γ^{jk},

$$| D_\lambda{}^s \frac{1}{P(\lambda, \xi + z\theta^{jk})} | \leqq M''(|s|) \frac{R(\xi)^{|s|}}{|P(\lambda, \xi + z\theta^{jk})|^{|s|+1}}$$
$$\leqq \frac{M''(|s|)}{b^{|s|+1}} \frac{R(\xi)^{|s|}}{\tilde{P}(\lambda,\xi)^{|s|+1}}.$$

With a new constant $M_1(|r|) < +\infty$ depending only on r, we get

$$\sum_k | D_\lambda{}^r F^{jk}(\lambda)| \leqq M_1(|r|) \int \beta^j(\xi) \left[\tau^j + \frac{R(\xi)}{\tilde{P}(\lambda,\xi)}\right]^{|r|} \frac{1}{\tilde{P}(\lambda,\xi)}$$
$$\times \sup_{|z|=1} | \hat{\phi}(\xi + z\theta^{jk}| \, d\xi.$$

Replacing τ^j by its definition, we see that

$$\sum_k | D_\lambda{}^r F^{jk}(\lambda)| \leqq M_2(|r|) \int \beta^j(\xi) \left[\frac{R(\xi^j)}{h(\xi^j)} + \frac{R(\xi)}{\tilde{P}(\lambda,\xi)}\right]^{|r|} \frac{1}{\tilde{P}(\lambda,\xi)}$$
$$\times \Big(\sup_{\theta \in C^n, |\theta| \leqq \epsilon} | \hat{\phi}(\xi + \theta)|\Big) d\xi.$$

We use then Lemma 4. We take into account Property i) of $R(\xi)$ and the fact that $\xi \in$ support of β^j implies $|\xi - \xi^j| < \eta$. We see thus that

$$\sum_k |D_\lambda{}^r F^{jk}(\lambda)| \leqq M_8(|r|) \int \beta^j(\xi)(1 + |\xi|)^{2m|r|+m} (\sup_{\theta \in C^n, |\theta| \leqq \epsilon} |\hat{\phi}(\xi + \theta)|)\, d\xi$$

$$\leqq M_3(|r|) \int \frac{\beta^j(\xi)}{(1 + |\xi|)^{n+1}} (1 + |\xi|^2)^{\bar{r}} (\sup_{\theta \in C^n, |\theta| \leqq \epsilon} |\hat{\phi}(\xi + \theta)|)\, d\xi$$

with $\bar{r} = 1 + m|r| +$ integral part of $\frac{1}{2}(m + n + 1)$. We obtain right away:

$$\sum_k |D_\lambda{}^r F^{jk}(\lambda)| \leqq M_4(|r|) \int \frac{\beta^j(\xi)\, d\xi}{(1 + |\xi|)^{n+1}} \sup_{\xi \in R^n} \sup_{\theta \in C^n, |\theta| \leqq \epsilon} [(1 + |\xi + \theta|)^{\bar{r}} \hat{\phi}(\xi + \theta)]$$

Let us set

$$N_\epsilon(\psi) = \int e^{\epsilon|x|} |\psi(x)|\, dx.$$

We have

$$\sup_{\xi \in R^n, \theta \in C^n, |\theta| \leqq \epsilon} |\hat{\psi}(\xi + \theta)| \leqq N_\epsilon(\psi).$$

Hence, calling Δ the Laplacian $\left(\frac{\partial}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial}{\partial x_n}\right)^2$,

$$\sum_k |D_\lambda{}^r F^{jk}(\lambda)| \leqq M_4(|r|) \int \frac{\beta^j(\xi)\, d\xi}{(1 + |\xi|)^{n+1}} N_\epsilon((1 - \Delta)^r \phi),$$

and therefore

$$\sum_{j,k} |D_\lambda{}^r F^{jk}(\lambda)| \leqq M_4(|r|) \int \frac{d\xi}{(1 + |\xi|)^{n+1}} N_\epsilon((1 - \Delta)^r \phi).$$

This shows precisely that

$$E(\lambda) = \frac{1}{(2\pi)^n} \sum_{j,k} E^{jk}(\lambda)$$

is a C^∞ function of $\lambda \in \frac{1}{2}\Lambda$ with values in the space of distributions $\mathcal{D}'_x$. That $E(\lambda)$ is a fundamental solution of $P(\lambda, D)$ is easy to check. For take $\phi(x) = P(\lambda, D_x)\psi(x)$. Then $\check{\phi} = P(\lambda, -D)\check{\psi}$ and

$$\langle E(\lambda), \check{\phi} \rangle = \langle P(\lambda, D) E(\lambda), \check{\psi} \rangle.$$

We have

$$\langle E^{jk}(\lambda), \check{\phi} \rangle = \int \gamma^{jk}(\xi, \lambda) \int_{|z|=1} \hat{\psi}(\xi + z\theta^{jk}) \frac{dz}{2i\pi z}\, d\xi = \int \gamma^{jk}(\xi, \lambda) \hat{\psi}(\xi)\, d\xi$$

since $\psi(\zeta)$ is an entire function of $\zeta \in C^n$. Thus:

$$\sum_{j,k} \langle E^{jk}(\lambda), \check{\phi} \rangle = \int \hat{\psi}(\xi)\, d\xi = (2\pi)^n \psi(0).$$

I. 5. Some local and global properties of $E(\lambda)$. We fix now λ in $\frac{1}{2}\Lambda$. Let $Q(D)$ be a differential polynomial weaker than $P(\lambda, D)$. Take $\phi = Q(D)u * v$, with $u, v \in C_0^\infty(R^n)$. We have $\check{\phi} = Q(-D)\check{u} * \check{v}$ and:

$$\langle E(\lambda), \check{\phi}\rangle = \langle Q(D)E(\lambda) * u, \check{v}\rangle.$$

On the other hand:

$$\langle E^{jk}(\lambda), \check{\phi}\rangle = \int \gamma^{jk}(\xi, \lambda) \int_{|z|=1} \frac{Q(\xi + z\theta^{jk})}{P(\lambda, \xi + z\theta^{jk})} \hat{u}(\xi + z\theta^{jk}) \hat{v}(\xi + z\theta^{jk}) \frac{dz}{2i\pi z}\, d\xi.$$

We have trivially

$$| Q(\xi + z\theta^{jk}) | \leqq C\tilde{Q}(\xi)$$

where $C < +\infty$ does not depend on $\xi \in \boldsymbol{R}^n$ nor on j, k. We use then (I. 9) and the fact that

$$\bar{Q}(\xi) \leqq C'\tilde{P}(\lambda, \xi), \quad \xi \in \boldsymbol{R}^n.$$

We recall that all the vectors θ^{jk} belong to the finite set S whose elements satisfy $|\theta| < \epsilon$. Setting $\epsilon_0 = \sup\limits_{j,k} |\theta^{jk}|$, we see that

$$| \langle E^{jk}(\lambda), \check{\phi}\rangle | \leqq C'' \int \gamma^{jk}(\xi, \lambda) \sup_{\theta \in \boldsymbol{C}^n, |\theta| \leqq \epsilon_0} | \hat{u}(\xi + \theta) \hat{v}(\xi + \theta) |\, d\xi,$$

whence

$$| \langle E(\lambda), \check{\phi}\rangle | \leqq C'' \int \sup_{\theta \in \boldsymbol{C}^n, |\theta| \leqq \epsilon_0} | \hat{u}(\xi + \theta) \hat{v}(\xi + \theta) |\, d\xi.$$

Note that $\epsilon_0 < \epsilon$. It can then be shown (cf. [2]) that

$$\int \sup_{\theta \in \boldsymbol{C}^n, |\theta| \leqq \epsilon_0} | \hat{u}(\xi + \theta) \hat{v}(\xi + \theta) |\, d\xi \leqq A \| e^{\epsilon |x|} u \|_{L^2} \| e^{\epsilon |x|} v \|_{L^2}.$$

Let us denote by $L^2(r)$ the space of measurable functions f such that $e^{r|x|} f \in L^2$; we provide $L^2(r)$ with the norm $\| e^{r|x|} f \|_{L^2}$. Then C_0^∞ is dense in $L^2(r)$ and the bilinear form $\int \phi\psi\, dx$ can be extended from $C_0^\infty \times C_0^\infty$ to $L^2(r) \times L^2(-r)$ and puts $L^2(r)$ and $L^2(-r)$ in duality; we may thus regard $L^2(-r)$ as the dual of $L^2(r)$. We have shown that

$$| \langle Q(D)E(\lambda) * u, \check{v}\rangle | \leqq A' \| e^{\epsilon |x|} u \|_{L^2} \| e^{\epsilon |x|} v \|_{L^2}.$$

This can be interpreted then as follows: $u \to Q(D)E(\lambda) * u$, defined on C_0^∞, can be extended as a bounded linear map from $L^2(\epsilon)$ into $L^2(-\epsilon)$. Since $L^2_c \subset L^2(\epsilon)$ and $L^2(-\epsilon) \subset L^2_{loc}$, this fact has the first consequence that $Q(D)E(\lambda) * L^2_c \subset L^2_{loc}$. In other words, $E(\lambda)$ is a *proper* fundamental solution of $P(\lambda, D_x)$ (cf. Hörmander [2]). It can also be proved that any distribution T such that $T * L^2(\epsilon) \subset L^2(-\epsilon)$ has the following additional property: $e^{-\epsilon(\sqrt{1+|x|^2})} T$ is a tempered distribution. We obtain in this way:

THEOREM 1′ *Same hypotheses as in Theorem* 1. *Let* $\epsilon > 0$ *be arbitrary. There is a* C^∞ *function* $\lambda \to E(\lambda)$ *of* Λ *into* $\mathcal{D}'_x$ *such that, for each* $\lambda \in \Lambda$,

1°) $E(\lambda)$ *is a proper fundamental solution of* $P(\lambda, D_x)$;

2°) $E(\lambda)$ $exp(-\epsilon\sqrt{1+|x|^2})$ *is a tempered distribution.*

Indeed, we have proved that for some locally finite open covering $\{\Lambda^\iota\}$ $(\iota \in I)$ of Λ, there exist fundamental solutions $E^\iota(\lambda)$ of $P(\lambda, D_x)$ when $\lambda \in \Lambda^\iota$ $(\iota \in I)$ which are proper and such that $E^\iota(\lambda) \exp(-\epsilon\sqrt{1+|x|^2})$ is tempered. If then $\{a^\iota(\lambda)\}$ $(\iota \in I)$ is a partition of unity in $C_0^\infty(\Lambda)$ subordinated to the covering $\{\Lambda^\iota\}$, the fundamental solution of $P(\lambda, D_x)$,

$$E(\lambda) = \sum_{\iota \in I} a^\iota(\lambda) E^\iota(\lambda)$$

enjoys obviously properties 1°) and 2°) for all $\lambda \in \Lambda$.

II. Proofs of Theorems 2 and 3.

II. 1. Dependence on the parameters. We keep assuming that Λ is an open ball in the euclidean space R^d and that the coefficients $A_p(\lambda)$ of $P(\lambda, D_x)$ are C^∞ functions and do not vanish simultaneously in an open ball $\Lambda' \subset R^d$ such that $\bar{\Lambda} \subset \Lambda'$. We make now the additional hypothesis:

(ES) *When* λ *varies in* Λ', *the differential operators* $P(\lambda, D_x)$ *are equally strong.*

Take $\lambda^0 \in \Lambda'$ arbitrary. Since the polynomials $P(\lambda, \xi)$, when λ runs over Λ', form a linear space of finite dimension, one can find a finite basis of that space including $P(\lambda^0, \xi)$; in other words we can write, for all $\lambda \in \Lambda'$,

$$P(\lambda, D_x) = a_0(\lambda) P(\lambda^0, D_x) + \sum_{j=1}^{r} a_j(\lambda) P_j(D_x) \tag{II.1}$$

where the differential polynomials $P_j(D)$ are necessarly weaker than $P(\lambda^0, D)$. The coefficients $a_j(\lambda)$ are (complex valued) C^∞ functions of $\lambda \in \Lambda'$.

LEMMA 5. *Assume that Property* (ES) *holds. There is a constant* $\gamma > 0$ *such that, for all* $\lambda, \lambda' \in \Lambda$, $\xi \in R^n$,

$$\tilde{P}(\lambda, \xi) \geqq \gamma \tilde{P}(\lambda', \xi). \tag{II.2}$$

Whatever be $\epsilon > 0$, there is a neighborhood $U(\epsilon)$ of λ^0 in Λ' such that for $\lambda \in U(\epsilon)$,

$$|a_0(\lambda) - 1| \leqq \epsilon, \; |a_j(\lambda)| \leqq \epsilon.$$

For a suitable constant $C < +\infty$, we have

$$\tilde{P}_j(\xi) \leqq C\tilde{P}(\lambda^0, \xi), \xi \in R^n, \qquad j = 1, \cdots, r.$$

For $\lambda \in U(\epsilon)$ we derive from (II.1)

$$\tilde{P}(\lambda, \xi) \geqq (1-\epsilon)\tilde{P}(\lambda^0, \xi) - C'\epsilon\tilde{P}(\lambda^0, \xi)$$

where $C' < +\infty$ does not depend on ξ, λ^0, ϵ. Thus for $\epsilon > 0$ small enough, $\lambda \in U(\epsilon)$ implies

$$\tilde{P}(\lambda, \xi) \geqq \tfrac{1}{2}\tilde{P}(\lambda^0, \xi), \xi \in R^n.$$

Since $\bar{\Lambda}$ is compact, we can find a finite set of points $\lambda^i \in \bar{\Lambda}$ $(i = 1, \cdots, I)$ such that, given any $\lambda \in \Lambda$, we have, for some i,

$$\tilde{P}(\lambda, \xi) \geqq \tfrac{1}{2}\tilde{P}(\lambda^i, \xi), \xi \in R^n.$$

Apply then once more (II.1) with $\lambda^0 = \lambda^i$ $(1 \leqq i \leqq I)$. We obtain:

$$\tilde{P}(\lambda', \xi) \leqq \sup_{\lambda' \in \bar{\Lambda}} \Big(\sum_{j=0}^{r} |a_j(\lambda')|\Big)\Big(\tilde{P}(\lambda^i, \xi) + \sum_{j=1}^{r} \tilde{P}(\xi)\Big)$$

There is a constant $C_i < \infty$, depending only on i, such that

$$\sum_{j=1}^{r} \tilde{P}_j(\xi) \leqq C_i\tilde{P}(\lambda^i, \xi), \xi \in R^n.$$

Of course, the number $M^i = \sup_{\lambda' \in \Lambda} \big(\sum_{j=0}^{r} |a_j(\lambda')|\big)$ depends also on $i = 1, \cdots, I$. Anyway, we obtain

$$\tilde{P}(\lambda', \xi) \leqq 2 \sup_{1 \leqq i \leqq I} [M^i(1 + C^i)] \cdot \tilde{P}(\lambda, \xi).$$

LEMMA 6. *For each multi-index* $r = (r_1, \cdots, r_d)$, *we have*

(II.3) $$|D_\lambda^r P^{(p)}(\lambda, \xi)| \leqq M(|r|)\tilde{P}(\lambda', \xi), \lambda, \lambda' \in \Lambda, \xi \in R^n,$$

where $M(|r|)$ *does not depend on* λ, λ', ξ *and* $p = (p_1, \cdots, p_n)$.

We have:

$$|D_\lambda^r P^{(p)}(\lambda, \xi)| \leqq \sum_{j=0}^{r} |D_\lambda^r a_j(\lambda)| \, |P_j^{(p)}(\xi)|,$$

setting $P_0(D) = P(\lambda^0, D)$. There is a constant $C < +\infty$ such that

$$\sum_{j=0}^{r} \tilde{P}_j(\xi) \leqq C\tilde{P}(\lambda^0, \xi), \xi \in R^n.$$

Applying then Lemma 5 yields the result.

The meaning of Lemma 6 is that, in Lemma 3, we can take as $R(\xi)$ any $\tilde{P}(\lambda', \xi)$, $\lambda' \in \Lambda$. Estimate (II.3) and the property of $M(|r|)$ shows that the constant $B(|r|)$ of Property (iii), § I.2, can be taken independently of λ', The same is true for the constants A, C in Properties i), ii) respectively, as easily shown. At this point, we use Remark I.1, § I.2: if we take $R(\xi) = \tilde{P}(\lambda', \xi)$, we may choose in Lemma 3 the constants η, a, b independently of λ'. By taking finally $\lambda' = \lambda^0$ we obtain:

LEMMA 3′. *Assume that Property* (ES) *holds. Then there are numbers* > 0, η, a, b *such that for every* $\xi^0 \in R^n$, $\lambda^0 \in \Lambda$, *one can find a vector* $\theta \in S$ *with the property that*

$$|\xi - \xi^0| \leqq \eta. \ |\lambda - \lambda^0| \leqq a, \ \lambda \in \Lambda,$$

implies

$$b\tilde{P}(\lambda, \xi) \leqq \inf_{|z|=1} |P(\lambda, \xi + z\theta)|.$$

At last we state the equivalent of Lemma 4:

LEMMA 4′. *There is a constant* $c > 0$ *such that for all* $\lambda \in \Lambda$, $\xi \in R^n$,

$$\tilde{P}(\lambda, \xi) \geqq c.$$

Trivial consequence of Lemma 5.

II. 2. Construction and properties of the fundamental solution. Under the assumption (ES), where therefore the Lemmas 3′ and 4′ are valid, we do not need to introduce the numbers τ^j (§ I.3). In fact, by restricting ourselves to the open ball $|\lambda| < a$, we may define the distributions $E^j(\lambda)$ by:

$$\langle E^j(\xi), \check{\phi} \rangle = \int_{|z|=1} \beta^j(\xi) \, \frac{\phi(\xi + z\theta^j)}{P(\lambda, \xi + z\theta^j)} \, \frac{dz}{2i\pi z} \, d\xi$$

where we have set $\theta^j = \theta^{j0}$.

Observe that if a differential polynomial $Q(D)$ is weaker than $P(\lambda, D)$ for some $\lambda \in \Lambda$, this will be true for all $\lambda \in \Lambda$. Moreover, in view of (II.2), $Q(D)$ is "uniformly weaker" than the $P(\lambda, D)$, that is to say

$$\tilde{Q}(\xi) \leqq C\tilde{P}(\lambda, \xi), \xi \in R^n, \tag{II.4}$$

where C does not depend on $\lambda \in \Lambda$.

Take then $\phi = Q(D)u * v$ with $u, v \in C_0{}^\infty(R^n)$. We have

$$\langle E^j(\lambda), \check{\phi} \rangle = \langle Q(D)E^j(\lambda) * u, \check{v} \rangle \ .$$

$$= \int \beta^j(\xi) \int \frac{Q(\xi + z\theta^j)}{P(\lambda, \xi + z\theta^j)} \hat{u}(\xi + z\theta^j) \hat{v}(\xi + z\theta^j) \frac{dz}{2i\pi z} d\xi.$$

Let us set $\Phi^j(\lambda) = \langle E^j(\lambda), \check{\phi}\rangle$. If $r = (r_1, \cdots, r_d)$, we have:

$$|D_\lambda{}^r\Phi^j(\lambda)| \leqq \int \beta^j(\xi) \int \left| D_\lambda{}^r \frac{1}{P(\lambda, \xi + z\theta^j)} \right| |Q(\xi + z\theta^j)| \, |\hat{u}(\xi + z\theta^j) \times \hat{v}(\xi + z\theta^j)| \frac{|dz|}{2\pi |z|} d\xi.$$

For $|\lambda| < a$ and $\xi \in$ support of β^j,

$$|P(\lambda, \xi + z\theta^j)| \geqq b\bar{P}(\lambda, \xi).$$

On the other hand, in view of (II. 3), there is a constant $C(r) < +\infty$ depending only on r such that

$$\left| D_\lambda{}^r \frac{1}{P(\lambda, \xi + z\theta^j)} \right| \leqq C(r) \frac{\bar{P}(\lambda, \xi)^{|r|}}{|P(\lambda, \xi + z\theta^j)|^{|r|+1}}.$$

Thus, with a different constant $C'(r)$, taking into account the fact that

$$|Q(\xi + z\theta^j)| \leqq A\bar{Q}(\xi) \leqq AC\bar{P}(\lambda, \xi)$$

(cf. (II. 4)), we get:

$$|D_\lambda{}^r\Phi^j(\lambda)| \leqq C'(r) \int \beta^j(\xi) \sup_{\theta \in C^n, |\theta| \leqq \epsilon_0} |\hat{u}(\xi + \theta)\hat{v}(\xi + \theta)| \, d\xi$$

and therefore

$$\text{(II. 5)} \qquad \sum_{j=0}^{\infty} |D_\lambda{}^r\Phi^j(\lambda)| \leqq C''(r) \int \sup_{\theta \in C^n, |\theta| \leqq \epsilon_0} |\hat{u}(\xi + \theta)\hat{v}(\xi + \theta)| \, d\xi.$$

Here $\epsilon_0 = \sup_j |\theta^j| < \epsilon$ (cf. § I. 5).

Let us take first $Q(D) = I =$ identity and $u(x) = \rho_\tau(x) = \tau^n\rho(\tau x)$ with $\rho \in C_0{}^\infty(R^n)$, $\int \rho(x)dx = 1$, $\rho \geqq 0$. By letting $\tau \to +\infty$, we obtain

$$\sum_{j=0} |D_\lambda{}^r \langle E^j(\lambda), \check{v}\rangle| \leqq C''(r) \int \sup_{\theta \in C^n, |\theta| \leqq \epsilon_0} |\hat{v}(\xi + \theta)| \, d\xi.$$

By an argument similar to the one of § I. 5, we conclude that

$$\text{(II. 6)} \qquad E(\lambda) = \frac{1}{(2\pi)^n} \sum_{j=0}^{\infty} E^j(\lambda)$$

is a C^∞ function of λ in the open ball $|\lambda| < a$, with values in $\mathcal{D}'$.

Then by the same argument as in § I. 5, we derive from (II. 5) that

$$|\langle [D_\lambda{}^r Q(D)E(\lambda)] * u, \check{v}\rangle| \leqq C_3(r) \| e^{\epsilon|x|}u \|_{L^2} \| e^{\epsilon|x|}v \|_{L^2}.$$

This shows that $\lambda \to Q(D)E(\lambda) *$ is a C^∞ function of λ, $|\lambda| < a$, with values in the space $\mathcal{L}(L^2(\epsilon); L^2(-\epsilon))$ of bounded linear operators $L^2(\epsilon) \to L^2(-\epsilon)$.

From the fact that $L^2_c \subset L^2(\epsilon)$ and $L^2(-\epsilon) \subset L^2_{loc}$, we obtain a strengthened form of Theorem 2:

THEOREM 2′. *Same hypotheses as in Theorem 2. Let* $\epsilon > 0$ *be arbitrary. There is a* C^∞ *function of* $\lambda \in \Lambda$, $E(\lambda)$, *with values in* $\mathcal{D}'_x$, *and with the following properties:*

1°) *for each* $\lambda \in \Lambda$, $E(\lambda)$ *is a proper fundamental solution of* $P(\lambda, D)$; *furthermore, if* $Q(D)$ *is any differential polynomial weaker than the* $P(\lambda, D)$, *then* $\lambda \to Q(D)E(\lambda) *$ *is a* C^∞ *function of* λ *with values in* $\mathcal{L}(L^2_c; L^2_{loc})$;

2°) $\lambda \to E(\lambda) exp(-\epsilon\sqrt{1+|x|^2})$ *is a* C^∞ *function in* Λ *with values in the space* $\mathcal{S}'_x$ *of tempered distributions.*

II. 3. The holomorphic case. Now the proof of Theorem 3 is trivial. Assume that Λ is a ball in a complex space C^d. Let us look at $E(\lambda)$ as defined in (II. 6) for $|\lambda| < a$. Whatever be $\phi \in C_0^\infty(R^n)$, $\langle E(\lambda), \check{\phi}\rangle$ is the sum of the uniformly absolutely convergent series $\frac{1}{(2\pi)^n}\sum_{j=0}^{\infty}\langle E^j(\lambda), \check{\phi}\rangle$. But under the hypothesis of Theorem 3,

$$\langle E^j(\lambda), \check{\phi}\rangle = \int \beta^j(\xi) \int_{|z|=1} \frac{\hat{\phi}(\xi + z\theta^j)}{P(\lambda, \xi + z\theta^j)} \frac{dz}{2i\pi z} d\xi$$

is obviously holomorphic in $|\lambda| < a$: $\langle E(\lambda), \check{\phi}\rangle$ must also be holomorphic in $|\lambda| < a$. In fact, the same reasoning applies to the operator $Q(D)E(\lambda) *$: $L^2(\epsilon) \to L^2(-\epsilon)$, where $Q(D)$ is weaker than the $P(\lambda, D)$. We conclude that $\lambda \to Q(D)E(\lambda) *$ is an holomorphic function in $|\lambda| < a$ with values in $\mathcal{L}(L^2(\epsilon); L^2(-\epsilon))$. From these follows the analogue of Theorem 2′:

THEOREM 3′. *Same hypotheses as in Theorem 3. Let* $\epsilon > 0$ *be arbitrary. Every point* λ^0 *of* Λ *has an open neighborhood* $N(\lambda^0)$ *in which is defined an holomorphic function* $E(\lambda)$, *valued in* $\mathcal{D}'_x$, *with the following properties:*

1°) *for each* $\lambda \in N(\lambda^0)$, $E(\lambda)$ *is a proper fundamental solution of* $P(\lambda, D)$; *furthermore, if* $Q(D)$ *is any differential polynomial weaker than the* $P(\lambda, D)$, *then* $\lambda \to Q(D)E(\lambda) *$ *is an holomorphic function of* $\lambda \in N(\lambda^0)$ *with values in* $\mathcal{L}(L^2_c; L^2_{loc})$;

2°) $\lambda \to E(\lambda) exp(-\epsilon\sqrt{1+|x|^2})$ *is an holomorphic function* λ *in* $N(\lambda^0)$ *with values in* $\mathcal{S}'_x$.

Theorem 3′ strengthens a result of Zerner [4].

III.

We may replace, in Theorem 1 or 1', everywhere C^∞ by C^ν $(1 \leqq \nu < \infty)$, or even by C^0 assuming then that Λ is a topological space (with suitable properties).[6] Let us assume here, for simplicity, that the order of $P(\lambda, D)$ is bounded by the integer $m \geqq 0$ when λ varies in Λ. For let

$$\mathfrak{P}(A, \xi) = \sum_{|p| \leqq m} A_p \xi^p$$

be the "general" polynomial in n variables, of degree m. We have set $A = (A_p)$; we may regard A as a parameter, varying in a complex space C^r, where $r = r(m, n)$ is the dimension of the linear space of all polynomials in n variables, of degree m. In order to get

$$P(\lambda, D) = \sum_{|p| \leqq m} A_p(\lambda) D^p, \ \lambda \in \Lambda,$$

one needs only to take $A = A(\lambda) = (A_p(\lambda))$, $\lambda \in \Lambda$, in $\mathfrak{P}(A, D)$. Theorem 1 states that there is a fundamental solution $E(A)$ of $\mathfrak{P}(A, D)$, for $A \in C^r - \{0\}$, which is a C^∞ function of $A \in C^r - \{0\}$. If then $\lambda \rightarrow A(\lambda)$ is a C^ν $(0 \leqq \nu < +\infty)$ function in λ, $E(A(\lambda))$, which is a fundamental solution of $P(\lambda, D)$, will also be a C^ν function in Λ.

REFERENCES.

[1] L. Hörmander, "On the theory of general partial differential operators," *Acta Mathematica*, vol. 94 (1955), p. 218.

[2] ———, "Local and global properties of fundamental solutions," *Mathematica Scandinavica*, vol. 5 (1957), p. 27.

[3] F. Treves, "Solutions élémentaires d'équations aux dérivées partielles dépendant d'un paramètre," *Comptes Rendus des Séances de l'Académie des Sciences* (Paris), vol. 242 (1956), p. 1250.

[4] M. Zerner, "Solution élémentaire locale d'équations aux dérivées partielles dépendant d'un paramètre," *Comptes Rendus des Séances de l'Académie des Sciences* (Paris), vol. 248 (1959), p. 3679.

---

[6] With C^0 we obtain thus a result similar to the one of Treves [3].

ON A CLASS OF INVERSE SEMIGROUPS.*

By P. S. VENKATESAN.

In this note we investigate the structure of a special class of inverse semigroups—inverse semigroups the non-zero idempotents of which are primitive. We show that an inverse semigroup S has its non-zero idempotents primitive if and only if it is a class sum of its Brandt ideals. A set of other equivalent conditions on S is also obtained. Brandt semigroups are treated by Mcfadden and Schneider [4] and Preston and Clifford [6]. The reader is referred to Bruck [1] and Preston and Clifford [6] for many results and literature on these subjects.

A *semigroup* is a non-null multiplicative associative system. Throughout S denotes a semigroup with zero. A subset A of S is a *left* (*right*) *ideal* of S if $SA \subseteq A$ $(AS \subseteq A)$. A is a *two-sided ideal* of S if it is both a left and a right ideal of S. A left ideal B of S is *minimal* if $B \neq (0)$ and if it contains no proper left ideal of S. Similarly minimal right (two-sided) ideal is defined. S is *simple* if it contains no proper two-sided ideal of S and $S^2 \neq (0)$. The *principal left* (*right*) *ideal* generated by an element a in S is $a \cup Sa$ $(a \cup aS)$.

S is said to be *regular* if for any a in S there exists x in S such that $a = axa$. $e = ax$ and $f = xa$ are idempotents and $ea = a = af$. $e(f)$ is called a *left* (*right*) *unit* of a. S is regular if and only if for any a in S, $a \cup aS = eS$ $(a \cup Sa = Sf)$ for some idempotent e (f) in S. A regular semigroup with commuting idempotents is *inverse*. An idempotent e in S is *primitive* if for any idempotent f in S, $ef = fe = f$ implies $f = 0$ or $f = e$. S is *completely simple* if S is simple and contains a non-zero primitive idempotent. S is a *Brandt semigroup* if:

(i) For each non-zero a in S three exist unique idempotents e, f in S such that $ea = a = af$.

(ii) For each pair of non-zero idempotents e, f in S there exists a non-zero a in S such that $ea = a = af$.

(iii) If $ea = a = af$ for non-zero a, e, f in S then there exists a unique b in S such that $ab = e$, $ba = f$ and $fb = b = be$.

(iv) For any a, b in S $ab \neq 0$ if and only if the right unit of a is the left unit of b.

---

* Received March 30, 1962.

We use the following known results.

(A) If R and L are minimal right and left ideals respectively of a semigroup S such that $LR \neq (0)$ and $RL \neq (0)$ then LR is a completely simple semigroup and $RL = R \cap L$ is a group with zero ([3], [7]).

(B) A semigroup S is a Brandt semigroup if and only if it is completely simple and inverse (p. 102 in [6]).

(C) If S is an inverse semigroup then for any a in S the equations $a = axa$ and $x = xax$ have a unique common solution [5].

1. Preliminary results.

LEMMA 1. *Let e be a non-zero idempotent in a semigroup S. Then e is primitive if and only if for any non-zero idempotent f in S, $ef = f$ implies $fe = e$.*

Proof. Now $ef = f$ implies $(efe)^2 = efe = fe$ and as $e(efe) = efe = (efe)e$ and $efe \neq 0$, $e = efe = fe$ because e is primitive. The converse is obvious.

COROLLARY. *If e, f are non-zero primitive idempotents in S then* (i) *$ef = f$ if and only if $fe = e$;* (ii) *$ef = e$ if and only if $fe = f$.*

LEMMA 2. (Ex. 10, p. 84 in [6]). *If e is a non-zero primitive idempotent in a regular semigroup S then eS (Se) is a minimal right (left) ideal of S.*

Proof. If eS is not minimal then let $RS \subseteq R \subseteq eS$. As S is regular, R contains a non-zero idempotent f. So $ef = f$ and by Lemma 1, $e = fe \in fS$. Thus $eS \subseteq fS \subseteq R \subseteq eS$ proving the minimality of eS. Similarly Se is minimal.

LEMMA 3. *If R and L are respectively right and left ideals of a regular semigroup S such that $LR \neq (0)$ then $RL \neq (0)$.*

Proof. If $LR \neq (0)$ then by regularity of S, LR contains a non-zero idempotent f. Now $0 \neq f = f^2 \in L(RL)R$ and hence $RL \neq (0)$.

LEMMA 4. *If every non-null principal right (left) ideal of a semigroup S is minimal and is generated by only one idempotent then S is inverse and the product of distinct idempotents in S is zero.*

Proof. Obviously S is regular. Let e, f be distinct non-zero idempotents in S with $ef \neq 0$. Then as the ideals fS and Se are minimal, by

Lemma 3 and (A) fSe is a group with zero. If g is the identity of the group, by minimality we have $gS = fS$ and $Se = Sg$ and so by hypothesis $e = g = f$ contradicting the assumption that e and f are distinct. Thus $ef = 0$ if e and f are distinct and S is inverse.

2. Inverse semigroups whose idempotents are primitive.

THEOREM 1. *A semigroup S with zero is an inverse semigroup in which every non-zero idempotent is primitive if and only if S is the union of its minimal ideals, each of which is a Brandt semigroup.*

Proof. Let S be an inverse semigroup in which every non-zero idempotent e_i $(i \in I)$ is primitive. Then as the idempotents commute, the product of distinct idempotents in S is zero. For any non-zero idempotent e_i in S, by (A) and Lemma 2, Se_iS is a completely simple semigroup and hence a minimal ideal of S. Clearly disinct Se_iS are disjoint. As S is inverse, each non-zero a in S is in some $e_iSe_j \subseteq Se_i = Se_jS$ and so S is a class sum of its minimal ideals Se_iS $(i \in I)$. By (B) and hypothesis on S, each of these minimal ideals is a Brandt semigroup. Hence we get the required result.

Conversely let S be the union of its minimal ideals, each of which is a Brandt semigroup. Clearly S is inverse. Let e, f be idempotents in S. If e and f both belong to the same minimal ideal then by (B) either $ef = 0$ or $e = f$. If e and f respectively belong to the minimal ideals S_α and S_β of S then $ef \in S_\alpha S_\beta \subseteq S_\alpha \cap S_\beta$. Now by minimality $S_\alpha S_\beta = (0)$ or $S_\alpha = S_\beta$ and so as before $ef = 0$ or $e = f$. Thus the product of distinct idempotents in S is zero and hence the non-zero idempotents in S are primitive as desired.

THEOREM 2. *The following statements on a semigroup S with zero are equivalent.*

(1) *S is a regular semigroup and for any non-zero idempotent e in S the equation $exe = e$ has no idempotent solution other than e.*

(2) *Every non-null principal right (left) ideal is minimal and is generated by only one idempotent.*

(3) *For each non-zero a in S there exists a unique x in S such that $a = axa$.*

(4) *For each non-zero a in S there exist unique idempotents e, f in S such that $ea = a = af$. And every non-null principal right (left) ideal contains just one non-zero idempotent.*

(5) *S is an inverse semigroup and the product of distinct idempotents in S is zero.*

(6) *S is an inverse semigroup and for any a, b, c, d in S, $0 \neq ab = ac$ implies $b = c$ (or equivalently $0 \neq ab = db$ implies $a = d$).*

Proof. (1) implies (2). As S is regular, for any non-zero a in S, $aS = eS$ for some idempotent e in S. Let R be a non-null right ideal of S contained in eS. If g is a non-zero idempotent in R, $eg = g$ and $geg = g$. By (1) $g = e$ and hence $gS = R = eS$ proving the minimality of eS and thus (2).

(2) implies (3). By Lemma 4, S is inverse. So for any non-zero a in S, $a = axa$ for some x in S. As x and xa are non-zero, by minimality $xaS = xS$. Thus the idempotent xa is a left unit for xS and since $x \in xS$, $xax = x$. But by (C) the equations $axa = a$ and $xax = x$ have a unique common solution and so x is the only element for which $a = axa$.

(3) implies (4). For any non-zero a in S there is an x in S such that $a = axa = ea = af$ where $0 \neq e^2 = e = ax$ and $0 \neq f^2 = f = xa$. If further $ga = a$ for some idempotent g in S then as $ge = e$, $ege = e$ and so by (3) $g = e$. Similarly f is unique. Again by (3) $aS = eS$ and eS contains only one non-zero idempotent e.

(4) implies (5). For any non-zero b in S, $bS = gbS \subseteq gS$ for some idempotent g in S. As bS contains just one non-zero idempotent it must only be g. So $bS = gS$ and every non-null principal right (and similarly principal left) ideal is minimal and is generated by only one non-zero idempotent. Therefore (4) implies (2) and hence the conclusion of Lemma 4. Thus (4) implies (5).

(5) implies (6). Let $x = ab = ac \neq 0$. From Theorem 1 and definition of Brandt semigroup, for the element a there exists a unique element a' such that $a'a = e$ and $ae = a$. Thus $a'x = eb = ec \neq 0$ which in turn implies $b = c$. Similarly $0 \neq ab = ad$ implies $a = d$.

(6) implies (1). Let e and f be a non-zero idempotents in S such that $efe = e$. Since $ef = fe$, we have $e = e^2 = efe = ef \neq 0$ and so by left cancellation we get $e = f$. Thus for any non-zero idempotent e in S the equation $exe = e$ has no idempotent solution other than e proving (1).

COROLLARY. *Let S' be a semigroup without zero. If for each a in S' there exists a unique x in S' such that $a = axa$, then S' is a group* [2].

I wish to thank Prof. V. S. Krishnan for his encouragement and Prof. A. H. Clifford for his suggestions and remarks.

UNIVERSITY OF MADRAS, INDIA.

REFERENCES.

[1] R. H. Bruck, *A survey of binary systems* (Berlin, 1958).

[2] R. Croisot, "Demi-groupe inversifs et demi-groupe reunions de demi-groupes simples," *Annales Scientifiques de l'École Normale Supérieure*, vol. 70 (1953), pp. 361-379.

[3] A. H. Clifford, "Semigroups without nilpotent ideals," *American Journal of Mathematics*, vol. 71 (1949), pp. 834-844.

[4] R. Mcfadden and Hans Schneider, "Completely simple and inverse semigroups," *Proceedings of the Cambridge Philosophical Society*, vol. 57 (1961), pp. 234-236.

[5] W. D. Munn and R. Penrose, "A note on inverse semigroups," *Proceedings of the Cambridge Philosophical Society*, vol. 57 (1955), pp. 396-399.

[6] G. B. Preston and A. H. Clifford, *The Algebraic Theory of Semigroups*, American Mathematical Society Surveys, 7 (1961).

[7] R. P. Rich, "Completely simple ideals of a semigroup," *American Journal of Mathematics*, vol. 71 (1949), pp. 883-885.

EACH DISK IN E^3 CONTAINS A TAME ARC.*

By R. H. Bing.[1]

An arc ab in E^3 is tame if there is a homeomorphim h of E^3 onto itself such that $h(ab)$ is a polygonal arc. The purpose of this note is to show that each topological disk in E^3 contains many tame arcs.

A 2-sphere S is tame if there is a homeomorphism of E^3 onto itself taking S onto a polyhedral 2-sphere. It is known that each arc on a tame 2-sphere is tame. We show that each disk contains tame arcs by showing that it shares arcs with a tame 2-sphere.

We recall that a Sierpinski curve is obtained by removing from a 2-sphere the interiors of a null sequence of mutually exclusive disks such that the sum of the disks is dense in the 2-sphere. (A null sequence is one the diameter of whose elements converge to zero.) Any set topologically equivalent to such a Sierpinski curve is called a Sierpinski curve. It may be shown that any two Sierpinski curves are homeomorphic.

One of the principal theorems of this note is the following.

THEOREM 1. *For each 2-sphere S in E^3 and each positive number ϵ there is a Sierpinski curve X in S such that*

1. *X lies in a tame 2-sphere, and*
2. *each component of $S-X$ is of diameter less than ϵ.*

Listing the above theorem first gives direction to this note. We digress to give some definitions and other theorems. These other theorems point to the proof of Theorem 1 and that proof is given after Theorem 7.

If A, B are homeomorphic sets we say

$$H(A,B) \leqq \epsilon$$

if there is a homeomorphism of A onto B that moves no point by more than ϵ.

We denote the distance function for a metric space by ρ. If f_1, f_2 are two mappings of a set A into a bounded metric space B, we use

$$\rho(f_1, f_2) = \text{least upper bound } \rho(f_1(x), f_2(x)), \quad x \in A.$$

An ϵ set is a set of diameter no more than ϵ.

---

* Received May 8, 1962.

[1] Work on this paper was supported by contract NSF-G11665.

We use interior in two senses but this seems unlikely to cause confusion. If S is a 2-sphere in E^3, we use $\operatorname{Int} S$ and $\operatorname{Ext} S$ to denote the bounded and unbounded components respectively of $E^3 - S$. If D is a disk, we let $\operatorname{Int} D$ be the difference of D and its combinatorial boundary.

If S is a 2-sphere in E^3 and ϵ is a positive number, we say that S can be ϵ approximated from both sides if there are 2-spheres S', S'' in $\operatorname{Int} S$, $\operatorname{Ext} S$ respectively such that $H(S,S') < \epsilon$, $H(S,S'') < \epsilon$.

Theorem 2 provides a clue as to how we show the 2-sphere mentioned in Condition 1 of Theorem 1 is tame. Theorem 2 is lifted from [2].

THEOREM 2. *A 2-sphere S in E^3 is tame if for each positive number ϵ, S can be ϵ approximated from both sides.*

If S is a 2-sphere, U is a component of $E^3 - S$, and $\epsilon > 0$, we say that *S can be ϵ almost approximated from U* provided there is a 2-sphere S' such that

1. $H(S,S') < \epsilon$ and
2. each component of $S' - U$ is an ϵ set.

We say that S can be almost approximated from both sides if for each positive number ϵ, it can be ϵ almost approximated from each of its complementary domains.

We make use of the following theorem lifted from [3]. It is proved by methods similar to those employed in [1].

THEOREM 3. *Each 2-sphere in E^3 can be almost approximated from both sides.*

The tame 2-sphere mentioned in Condition 1 of Theorem 1 is obtained by a sequence of adjustments of the 2-sphere S of Theorem 1. Iterated use of an extension of the following result is made in getting these adjustments.

THEOREM 4. *For each 2-sphere S in E^3 and each positive number ϵ there is a 2-sphere S' such that*

1. *S' is obtained by removing a finite collection of mutually exclusive ϵ disks from S and replacing them with other ϵ disks, and*
2. *S' can be ϵ approximated from its interior.*

Proof of Theorem 4. We give the proof of Theorem 4 in seven steps.

1. Let δ be a positive number such that each δ set in S lies in a disk in S of diameter less than $\epsilon/6$.

2. It follows from Theorem 3 that there is a polyhedral 2-sphere S'' such that S'' contains a finite collection of mutually exclusive δ disks $D_1, D_2, \cdots, D_n$ such that

$$S'' - \sum \operatorname{Int} D_i \subset \operatorname{Int} S \text{ and } H(S, S'') \leqq \epsilon/6.$$

We note that each component of $S \cdot S''$ is a δ set.

3. We suppose that diameter $S > 2\epsilon$ and that there is a point p of $\operatorname{Int} S$ whose distance from S is more than ϵ. Let $E_1, E_2, \cdots, E_m$ be a finite collection of mutually exclusive $\epsilon/6$ disks in S such that

$$S \cdot S'' \subset \sum \operatorname{Int} E_j.$$

To see that there are such disks, consider the decomposition space whose nondegenerate elements are components of $S \cdot S''$. Then consider the big 2-sphere of the resulting cactoid [5]. A 0-dimensional set on this 2-sphere comes from $S \cdot S''$ and this closed 0-dimensional set can be covered with a finite collection of interiors of mutually exclusive very small disks on the 2-sphere. The boundaries of these disks in the 2-sphere of the cactoid correspond to the boundaries of the E_i's in S. The above reasoning is given in more expanded form in Theorem 9 of [3].

We note that

$$S - \sum E_i \subset \operatorname{Ext} S''$$

To see that this true, let qp be an arc which misses $\sum D_i + \sum E_i$ such that $q \in \operatorname{Ext} S$, $p \in \operatorname{Int} S$, and $\rho(\{p\} + \{q\}, S) > \epsilon$. It follows from Theorem 14 of [3] that there is such an arc. The first point of $(S - \sum E_i) + (S'' - \sum D_i)$ on qp in the order from q to p is a point of S since $S'' - \sum D_i \subset \operatorname{Int} S$.

4. It follows from Theorem 7 of [1] that each E_i can be replaced by a disk F_i such that

F_i is locally polyhedral at each point of $\operatorname{Int} F_i$,
diameter $F_i < \epsilon/4$, and
$(S - \sum E_i) + \sum F_i = S_1$ is a 2-sphere that contains $S'' - \sum \operatorname{Int} D_i$ on its interior.

5. With no loss of generality we suppose that each D_i and each F_j are in general position in the sense that each component of $D_i \cdot F_j$ is a polygonal simple closed curve. Such general position may be achieved by insuring that no four vertices of the triangulations of the D_i's and the F_j's are coplanar. Let U be the component of $S_1 \cdot \operatorname{Ext} S''$ containing $S - \sum E_i$. We note that

it is obtained by subtracting a finite collection of mutually exclusive disks from S_1. Each of these subtracted disks has its boundary on a D_i.

6. The 2-sphere S' is obtained as follows. Each component of $S_1 - U$ is a disk with its boundary on a D_i. Enlarge this disk slightly so that its boundary no longer intersects this D_i and replace the enlarged disk by an $\epsilon/6$ disk near this D_i and in $\operatorname{Ext} S''$. The holes in U may thus be enlarged and filled in one at a time. So that the resulting "fill-ins" will not intersect each other it may be convenient to plug up the small holes first by starting with a component J of $\operatorname{Bd} U$ on a D_i such that $\operatorname{Int} J$ on D_i does not intersect $\bar{U}$. Each hole in U may be thus treated so as to obtain a 2-sphere S' so that

$$S'' \subset \operatorname{Int} S', \text{ and } S - \sum E_i \subset S'.$$

Each component of $S' - (S - \sum E_i)$ is of diameter less than $\epsilon/4 + 2\epsilon/6 = 7\epsilon/12$.

7. Since the disks E_i have diameter less than $\epsilon/5$ and the disks of $S' - (S - \sum \operatorname{Int} E_i)$ have diameter less than $7\epsilon/12$,

$$H(S, S') < \epsilon/5 + 7\epsilon/12 < 5\epsilon/6.$$

This result with the inequality $H(S, S'') < \epsilon/6$ implies that

$$H(S', S'') < \epsilon.$$

The following is a slight extension of Theorem 4.

THEOREM 5. *For each 2-sphere S and each positive number ϵ there is a 2-sphere S' such that*

1. *S' is obtained by subtracting a finite collection of mutually exclusive ϵ disks from S and replacing them with ϵ disks, and*
2. *S' can be ϵ approximated from its interior and its exterior.*

Proof of Theorem 5. Apply Theorem 4 to get a 2-sphere S_1 such that

1. S_1 is obtained by subtracting a finite collection G of $\epsilon/4$ disks from S and replacing them with $\epsilon/4$ disks, and
2. there is a 2-sphere S'' on the interior of S_1 such that $H(S_1, S'') < \epsilon/4$.

Pick a small δ. (We indicate how small after the following paragraph.)

Reapply Theorem 4 (replacing $\operatorname{Int} S$ of the statement of that theorem by $\operatorname{Ext} S_1$) to get a 2-sphere S' such that

1. S' is obtained by subtracting a finite number of mutually exclusive δ disks from S_1 and replacing them with δ disks, and
2. there is a 2-sphere S''' on $\operatorname{Ext} S'$ such that $H(S', S''') < \delta$.

If we have picked δ small enough, S' is the required 2-sphere.

Finally, we indicate how small to take δ. We put on 4 restrictions.

Restriction 1. We take $\delta < \epsilon/4$. (We make more stringent restrictions later but this makes for an easy start.) This insures that $H(S_1, S') < \epsilon/2$ and $H(S', S'') < \epsilon$.

Restriction 2. We take δ so small that S'' lies on the interior of any 2δ approximation of S_1. This insures that $S'' \subset \operatorname{Int} S'$.

Restriction 3. We take δ so small that if A is a closed set in S each component of which is a δ set, then there is a finite collection H of mutually exclusive $\epsilon/3$ disks in S such that each point of A lies on the interior of one of these disks as does each element of G. We let A be the intersection of S and the δ disks of S_4 which are given by the second application of Theorem 4. The resulting collection H is the collection of disks in S to be replaced as suggested in the first condition of the statement of Theorem 5.

Restriction 4. Finally we take $\delta < \epsilon/12$. The purpose of this restriction is to insure that the disks in S' replacing the elements of H have diameters less than ϵ. Since each element of H is of diameter less than $\epsilon/3$ and the disks in S_1 replacing elements of G have diameters less than $\epsilon/4$, the small disks in S_1 sharing boundaries with disks of H have diameters less than $\epsilon/3 + \epsilon/4 + \epsilon/4 = 5\epsilon/6$. Since the disks in S' replacing disks in S_1 have diameters less than δ, the small disks of S' sharing boundaries with disks of H have diameters less than $5\epsilon/6 + 2\delta$, so if $\delta < \epsilon/12$, the disks in S' replacing the disks of H have diameters less than ϵ.

Another theorem we mention before proving Theorem 1 is the following.

THEOREM 6. *Suppose* $\epsilon > 0$ *and* $X_1, X_2, \cdots$ *is a sequence of compact sets in a metric space such that*

1. *each* X_i *has only a finite number of components,*
2. *each component of* X_1 *is an* $\epsilon/2$ *set, and*
3. *each component of* X_{i+1} $(i \geqq 1)$ *is of diameter less than* $e/2^{i+1}$ *and* $1/2^{i+1}$ *times the distance between any two components of*

$$X_1 + X_2 + \cdots + X_i.$$

Then there is a countable collection of mutually exclusive closed connected sets $A_1, A_2, \cdots$ *such that*

1. *each* A_k *is of diameter less than* ϵ,

2. $\lim (i \to \infty)$ *diameter* $A_i = 0$, *and*

3. *each component of each* X_i *lies in some* A_j.

For each point p in X_i, let $C(j,p)$ $(j > i)$ be the component of $X_1 + X_2 + \cdots + X_j$ containing p. The A_k containing p is the closure of $C(i,p) + C(i+1,p) + \cdots$. The proof that these A's satisfy the conclusion of the theorem is straightforward and we do not include it in this note.

We need the following result giving conditions under which the limit of a sequence of homeomorphisms is a homeomorphism

THEOREM 7. *Suppose* $f_1, f_2, \cdots$ *is a sequence of homeomorphisms of a compact metric space* X *into a complete metric space* Y *and* $\epsilon_1, \epsilon_2 \cdots$ *is a sequence of positive numbers such that* $\rho(f_i, f_{i+1}) < \epsilon_i$. *Then* $\lim f_i(X)$ *is a homeomorphism if* $\sum_j^\infty \epsilon_i$ *is a positive number so small that if* $\rho(x_1, x_2) > 1/j$ $(x_1, x_2 \in X)$, *then* $\rho(f_j(x_1), f_j(x_2)) > 2 \sum_j^\infty \epsilon_i$.

Again the proof is straightforward and we do not include it.

We finally give the proof of Theorem 1.

Proof of Theorem 1. The tame 2-sphere that is to share a Sierpinski curve with S is the limit of a sequence of 2-spheres $S_1, S_2, \cdots$. We indicate restrictions to be placed on the S_i's to insure that

1. $\lim S_i$ is a 2-sphere,

2. $\lim S_i$ is tame, and

3. $\lim S_i$ shares a Sierpinski curve X with S.

Step 1. *Restrictions to make* $\lim S_i$ *a 2-sphere.* Each S_i is the image of S under a homeomorphism h_i where $\rho(h_{i+1}, h_i) < \epsilon_i$ and $\epsilon_1, \epsilon_2, \cdots$ is a sequence of positive numbers chosen according to the criteria of Theorem 7 to make $\lim h_i$ a homeomorphism. Although in later steps we place more restrictions on $\epsilon_1, \epsilon_2, \cdots$, it follows from Theorem 7 that $\lim h_i$ is a homeomorphism if the ϵ_i's satisfy the following conditions. Insist (along with other restrictions to be imposed later) that $\epsilon_1 < 1/6$. After h_1 is defined (so that it moves no point more than ϵ_1), pick a positive number δ_1 so small that if x_1, x_2 are two points of S_1 such that $\rho(x_1, x_2) > 1/2$, then $\rho(h_1(x_1), h_1(x_2) > \delta_1$. Insist (among other things) that ϵ_2 is less than either

$\epsilon_1/2$ or $\delta_1/6$. After h_2 is defined (so that $\rho(h_1, h_2) < \epsilon_2$), pick a positive number δ_2 such that if x_1, x_2 are points of S_1 such that $\rho(x_1, x_2) > 1/3$, then $\rho(h_2(x_1), h_2(x_2)) > \delta_2$. Insist that $\epsilon_3 < \epsilon_2/2$, $\epsilon_3 < \delta_2/6$. This procedure is continued. We note that $\sum \epsilon_i < 1/3$ and $\sum_{j+1}^{\infty} \epsilon_i < \delta_j/3$ where δ_j is so small that if $\rho(x_1, x_2) > 1/j$, then $\rho(h_j(x_1), h_j(x_2)) > \delta_j$. Hence

$$\rho(h_{j+1}(x_1), h_{j+1}(x_2)) > \delta_j - 2\epsilon_{j+1} > 2\delta_{j/3} > 2\sum_{j+1}^{\infty} \epsilon_i.$$

Step 2. *Restrictions to make* $\lim S_i$ *tame.* Theorem 2 and Condition 2 of Theorem 5 provide the clues. We note from Theorem 5 that there is a 2-sphere S_1 and a homeomorphism h_1 of S onto S_1 such that h_1 moves no point by more than ϵ_1 and S_1 can be ϵ_1 approximated from $\operatorname{Int} S_1$ and $\operatorname{Ext} S_1$ by 2-spheres S_1', S_1'' respectively We pick $\sum_2^{\infty} \epsilon_i$ to be so small that S_1', S_1'' lie in the interior and exterior respectively of $\lim S_i = \lim h_i(S)$. In general, S_j is chosen so that it can be ϵ_j approximated by 2-spheres S_j', S_j'' in $\operatorname{Int} S_j$, $\operatorname{Ext} S_j$ respectively and $\sum_j^{\infty} \epsilon_i$ is chosen so small that S_j', S_j'' lie respectively in the interior and exterior of $\lim S_i = \lim h_i(S)$. Theorem 2 insures that $\lim S_i = \lim h_i(S)$ is tame.

Step 3. *Restrictions to make* $S \cdot \lim S_i$ *contain a Sierpinski curve* X. We use Theorem 6 and Condition 1 of Theorem 5. Let δ be a positive number such that if K is a δ set in S, $S - K$ has only one component of diameter as much as $\epsilon/2$. The number δ is to correspond to the ϵ of Theorem 6 so we pick $\epsilon_1 < \delta/2$. We find from Condition 1 of Theorem 5 that we can pick S_1 so that it is obtained by removing a finite collection H_1 of mutually exclusive ϵ_1 disks from S and replacing them with other ϵ_1 disks. The sum of the disks of H_1 corresponds to the set X_1 of Theorem 6.

We pick ϵ_2 less than either $\epsilon_1/4$ or $1/4$ the distance between any two elements of H_1 and choose S_2 so that it is obtained by removing a finite collection H_2 of mutually exclusive ϵ_2 disks from S_1 and replacing them with ϵ_2 disks. The set X_2 is a subset of S with only a finite number of components (each of diameter less than ϵ_2) such that X_2 contains the intersection of S with the sum of the elements of H_2. If this pattern is continued to get $S_3, S_4, \cdots$, Theorem 6 assures us that there is a null sequence $A_1, A_2, \cdots$, of mutually exclusive continua in S such that each A_j is a δ set and $S - \sum A_i$ lies in each S_i and hence in $S \cdot \lim S_i$.

The sum B_i of A_i and all small components of $S - A_i$ is an ϵ set. If one considers the upper semicontinuous decomposition G of S whose nondegenerate elements are the maximal B_i's, one finds from a theorem due to R. L. Moore [5] that the corresponding decomposition space is a 2-sphere. Since there are only a countable number of nondegenerate elements in G, one can find in this decomposition space a Sierpinski curve missing the points corresponding to the nondegenerate elements of G such that the corresponding Sierpinski curve X in S satisfies Condition 2 of Theorem 1.

The following result was announced at the International Congress of Mathematicians in Edinburgh in 1958. See Theorem 5 of [4].

THEOREM 8. *For each interior point p of each disk D in E^3 there is a 2-sphere S that contains an open subset of D containing p.*

A combination of Theorems 1 and 8 gives the following result.

THEOREM 9. *Each disk in E^3 contains many tame arcs.*

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] R. H. Bing, "Approximating surfaces with polyhedral ones," *Annals of Mathematics*, vol. 65 (1957), pp. 456-483.

[2] ———, "Conditions under which a surface in E^3 is tame," *Fundamenta Mathematicae*, vol. 47 (1959), pp. 105-139.

[3] ———, "Approximating surfaces from the side," *Annals of Mathematics*, vol. 77 (1963), in press.

[4] ———, "A surface is tame if its complement is 1-ULC," *Transactions of the American Mathematical Society*, vol. 101 (1961), pp. 456-483.

[5] R. L. Moore, "Concerning upper semicontinuous collections of continua," *Transactions of American Mathematical Society*, vol. 27 (1925), pp. 416-428.

EACH DISK IN E^3 IS PIERCED BY A TAME ARC.*

By R. H. Bing.[1]

1. Introduction.

1. Introduction. An arc ab in E^3 is *tame* if there is a homeomorphism of E^3 onto itself that takes ab onto a straight line interval. If ab lies in a plane in E^3, it is tame since if ab is an arc in a plane, there is a homeomorphism of the plane onto itself taking ab onto a straight line interval; this homeomorphism may be extended to E^3.

A disk is *pierced by* an arc ab at a point p if $p \in \operatorname{Int} D$, $p \in \operatorname{Int} ab$, p is the only point of $D \cdot ab$, and the arcs ap, pb abut on D from opposite sides in that there is an $\epsilon > 0$ such that no connected ϵ-set in $E^3 - D$ intersects both ap and pb. (An ϵ-set is a set of diameter less than or equal to ϵ.) If D lies on a 2-sphere which intersects ab only at p, the 2-sphere separates a from b in E^3 since each complementary domain of a 2-sphere in E^3 is uniformly locally connected. For convenience we state this result as a theorem.

THEOREM 1.1. *Suppose in E^3 that D is a disk that lies on a 2-sphere S and ab is an arc which intersects S only at a point p of* $\operatorname{Int} D$. *Then ab pierces D if and only if S separtes a from b.*

It is sometimes more convenient to study the interior and exterior of a 2-sphere in E^3 than to study the sides of disk. We show in Theorem 2.1 that this approach may be used since for each interior point p of a disk D in E^3, there is a disk D' in D such that $\operatorname{Int} D'$ contains p and lies on a 2-sphere.

Bing has given an example [1] of an arc in E^3 that pierces no disk. Fox and Artin have given an example [5] of a 2-sphere S with a point p such that no tame arc in E^3 pierces S at p. This 2-sphere S has something of a knotted feeler. By using a null sequence of knotted feelers one can construct a 2-sphere that cannot be pierced at any point of a countable dense set.

One might wonder if there is a disk so wild that it is pierced by no time arc. See the question on page 343 of [1]. One of the principal results

---

* Received May 8, 1962.

[1] Work on this paper was supported by contract NSF-G11665.

of this paper is to show in Theorem 5.3 that each disk in E^3 can be pierced by a tame arc.

In showing that each disk in E^3 can be pierced by a tame arc, we make extensive use of the fact (Theorem 1 of [4]) that each 2-sphere contains a Sierpinski curve X that lies on a tame 2-sphere. Section 3 of our paper is devoted to a study of Sierpinski curves.

2. Extending disks to 2-spheres. An arc A in the plane locally has two sides and we speak of arcs abutting on A either from the same side or from opposite sides. To aid us in studying the sides of A we may use either of two approaches. (1) We may use the fact that there is a homeomorphism of the plane onto itself taking A onto a horizontal segment and use the elementary notions of "above" and "below." (2) We may use the fact that A lies on a simple closed curve J and use the notion of the interior and exterior of J.

Neither of these approaches works with great ease in studying the sides of a disk in E^3. That the first fails may be seen by considering a wild disk D in E^3 such that there is no homeomorphism of E^3 onto itself taking D onto a polyhedral disk. The second is not completely successful since there are disks in E^3 which are not a subset of any 2-sphere. The disk shown in Figure 1 has a boundary which cannot be shrunk to a point in the complement of the disk. However, the second approach is partially successful as a result of the following theorem.

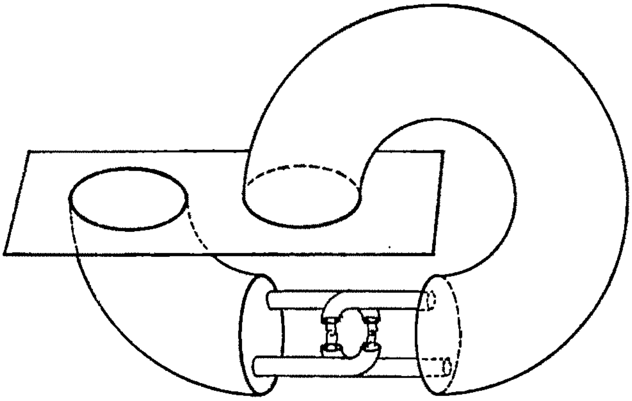

FIGURE 1.

THEOREM 2.1. *Suppose D is a disk in E^3, $p \in \operatorname{Int} D$, and $\epsilon > 0$. Then there is an ϵ 2-sphere S in E^3 and a disk D' in D such that $p \in \operatorname{Int} D' \subset S$.*

Theorem 2.1 is Theorem 5 of [3] but it is included here for completeness.

Proof. Let C be a cube in E of diameter less than ϵ and such that $p \in \operatorname{Int} C$, $\operatorname{Bd} D \subset \operatorname{Ext} C$.

Let f be a nonnegative continuous function defined on E^3 such that $f = 0$ in a neighborhood of p, f is positive on $\operatorname{Bd} C$, and $f < \rho(C, \operatorname{Bd} D)$.

It follows from Theorem 7 of [2] that there is a homeomorphism h of D into E^3 such that for each point x of D, $\rho(x, h(x)) \leqq f(x)$ and $h(D)$ is locally polyhedral at $h(x)$ if $f(x) > 0$. With no loss of generality we suppose $h(D)$ and $\operatorname{Bd} C$ are in general position.

Let K be the component of $h(D) \cdot C$ that contains p. It is a punctured disk and has as a boundary the sum of a finite collection of mutually exclusive simple closed curves $J_1, J_2, \cdots, J_n$. Let $E_1, E_2, \cdots, E_n$ be a finite collection of mutually exclusive disks in $\operatorname{Ext} C + \operatorname{Bd} C$ such that $\operatorname{Bd} E_i = J_i$ and diameter $(C + \sum E_i) < \epsilon$. The 2-sphere S promised by the theorem is $K + \sum E_i$.

3. Properties of Sierpinski curves.

In this section we review some properties of Sierpinski curves. We recall that a Sierpinski curve is the set obtained by removing from a 2-sphere S the interiors of a null sequence of mutually exclusive disks $\{D_i\}$ in S such that $\sum D_i$ is dense in S. The Sierpinski curve is

$$X = S - \sum \operatorname{Int} D_i.$$

Any two Sierpinski curves described in such a fashion are homeomorphic and any continuum homeomorphic to such a Sierpinski is called a Sierpinski curve irrespective of whether or not it is imbedded in a 2-sphere.

A convenient way to learn the properties of $X = S - \sum \operatorname{Int} D_i$ is to note that the decomposition G of $S - \operatorname{Int} D_1$ whose elements are the disks of $\{D_i\}$; $i = 2, 3, \cdots$ and the points of $(S - \sum D_i) + \operatorname{Bd} D_1$ is upper semicontinuous and use the result of R. L. Moore [7] that there is a monotone map f of $S - \operatorname{Int} D_1$ onto a disk D taking elements of the decomposition G onto unique points of D.

By considering such a decomposition of S one may obtain the following result.

THEOREM 3.1. *Suppose X is a Sierpinski curve on a 2-sphere S. For* *each $\epsilon > 0$ and each point p of X that is not a boundary point of a component of $S - X$ there is an ϵ disk D in S such that*

$$p \in \operatorname{Int} D, \qquad \operatorname{Bd} D \subset X.$$

A point of $X = S - \sum \operatorname{Int} D_i$ is an *accessible point* of X if it is a boundary point of some D_i; otherwise it is called *inaccessible*. Each such accessible point lies on an arc in X whose complement in X is uniformly locally connected; no inaccessible point lies on such an arc. Since lying on such an arc is a topological property, we have the following result.

THEOREM 3.2. *Suppose X is a Sierpinski curve on a 2-sphere S and h is a homeomorphism of X into a 2-sphere S'. The point p of X is a boundary point of a component of $S - X$ if and only if $h(p)$ is a boundary point of a component of $S' - h(X)$.*

A simple closed curve J is said to pierce a 2-manifold M at a point p if J contains an arc ab and M a disk D such that ab pierces D at p. The following definition of linking is a modification of the one given in [2]. The simple closed curve J_1 links the simple closed curve J_2 if J_1 bounds a 2-manifold M such that J_2 intersects M in an odd number of points and J_2 pierces M at each point that it intersects it. Furthermore, if J_1 links J_2 in this fashion and J_1', J_2' are closed curves (simple or singular) such that J_1' is homotopic to J_1 in $E^3 - J_2'$ and J_2' is homotopic to J_2 in $E^3 - J_1$, then J_1' is said to link J_2'. It was shown in [2] that if J_1' links J_2', then J_2' links J_1'. Hence, if J_1' links J_2', we say that J_1' and J_2' link.

Although the following theorem does not treat Sierpinski curves, it is used in Theorem 3.4 which does.

THEOREM 3.3. *Suppose S is a 2-sphere in E^3, J is a simple closed curve in S that bounds a disk D in S, and apb, aqb are arcs in E^3 such that*

$$a,\ b \text{ belong to different components of } E^3 - S,$$
$$apb \cdot (S - D) = 0,$$
$$aqb \cdot D = 0.$$

Then $apb + aqb$ and J link.

Proof. Let h be a homeomorphism of S onto a polyhedral 2-sphere in E^3 such that

$h(S)$ separates a from b,
$apb \cdot h(\overline{S-D}) = 0$,
$aqb \cdot h(D) = 0$,
there is a homotopy taking J to $h(J)$ in $E^3 - (apb + aqb)$.

That there is such a homeomorphism close to the identity follows from Theorem 1 of [2] and that $h(S)$ separates a from b under such a homeomorphism follows from Theorem VI 10 on page 97 of [6].

Let f be a homeomorphism of $apb + aqb$ onto a polygon in E^3 such that

f is the identity on a and b.
$f(apb) \cdot h(\overline{S-D}) = 0$,
$f(aqb) \cdot h(d) = 0$,
there is a homotopy taking $apb + aqb$ to $f(apb + aqb)$ in $E^3 - h(J)$.

With no loss of generality we suppose $f(apb)$ and $h(D)$ are in general position.

Then $f(apb + aqb)$ links $h(J)$ since $f(apb)$ crosses $h(D)$ an odd number of times and $f(aqb) \cdot h(D) = 0$. See page 480 of [2]. Since there is a homotopy taking $(apb + aqb)$ to $f(apb + aqb)$ in $E^3 - h(J)$, it follows from Theorem 10 of [2] that $apb + aqb$ links $h(J)$. Since there is a homotopy taking J to $h(J)$ in $E^3 - (apb + aqb)$, it follows from this same theorem that $apq + aqb$ links J.

THEOREM 3.4. *Suppose X is a Sierpinski curve that lies in each of two disks D_1, D_2 in E^3 and apb is an arc that intersects $D_1 + D_2$ only at an inaccessible point p of X. Then apb pierces D_1 if and only if it pierces D_2.*

Proof. Let D_i' be a subdisk of D_i such that D_i' lies in a 2-sphere S_i and p is in interior point of D_i'. That there is such a disk D_i' and a 2-sphere S_i follows from Theorem 2.1. Then apb pierces D_i if and only if it pierces D_i'. With no loss of generality we suppose that apb is so short that it intersects $S_1 + S_2$ only at p.

It follows from Theorem 1.1 that apb pierces D_i if and only if a, b belong to different components of $E^3 - S_i$. Suppose a, b belong to one component of $E^3 - S_1$. We show in the next two paragraphs that in this case they also belong to the same component of $E^3 - S_2$.

Let aqb be an arc from a to b in $E^3 - S_1$ such that $apb + aqb$ is a simple closed curve. Since $apb + aqb$ intersects S_1 only at p, any simple closed curve in $(X - p) \cdot D_1'$ can be shrunk to a point in $S_1 - (apb + aqb)$. Hence no simple closed curve in $(X - p) \cdot D_1'$ links $apb + aqb$. See Theorem 10 of [2].

Let J be a simple closed curve in $X \cdot D_2'$ which bounds a disk E in D_2' such that $p \in \operatorname{Int} E$ and diameter E is less than the distance between p and $aqb + (D_1 - D_1')$. That there is such a disk E follows from Theorem 3.1. Then a, b belong to the same component of $E^3 - S_2$ because if they did not, it would follow from Theorem 3.3 that J would link $apb + aqb$. Thus would contradict the conclusion of the preceding paragraph.

We have shown that if apb does not pierce D_1, then it does not pierce D_2. Similarly, it may be shown if it does not pierce D_2, it does not pierce D_1.

4. Arcs in disks.

The tame arcs that we will show pierce an arbitrary 2-sphere in E^3 lie in tame disks. In this section we study how arcs in disks may miss certain closed sets. An arc xy is said to span a disk D if the ends of xy lie on $\operatorname{Bd} D$ and xy lies except for its ends in $\operatorname{Int} D$.

THEOREM 4.1. *Suppose D is a disk, xy is a spanning arc of D, and Y is a closed subset of D containing xy such that for each $\epsilon > 0$ and each point q of* $\operatorname{Int} xy$ *there is a neighborhood N of q that intersects no component of $Y - xy$ of diameter greater than ϵ. Then there are points a, b of D on opposite sides of xy such that for each point p of xy there is an arc apb in $(D - Y) + \{p\}$ from a to b.*

Proof. Let E be a subdisk of D such that xy spans E and no component of $(Y - xy) \cdot E$ has either x or y as a limit point. The disk E is obtained by removing parts of D on both sides of xy near x and y without removing any point of xy. With no loss of generality we suppose that $D = E$.

Let C be the collection of components of Y other than xy and C' be the collection of all continua which are the sum of an element c of C and all components of $D - c$ which fail to contain xy. Ig G is the decomposition of D whose nondegenerate elements are the maximal elements of C', G is upper semicontinuous. Since no element of G separates D, it follows from a result of R. L. Moore [7] that there is a map f of D onto itself that takes each element of G onto a point of D and no two elements of G onto the same point. We suppose with no loss of generality that f is the identity on xy.

Since the set $f(Y)-xy$ is 0-dimensional, there is a null sequence Z of mutually exclusive disks in $D-xy$ which cover $f(Y)-xy$ such that each point of $f(Y)-(xy+\mathrm{Bd}\,D)$ lies on the interior of one of these disks and each point of $f(Y)-xy$ on $\mathrm{Bd}\,D$ lies on an open arc on $\mathrm{Bd}\,D$ which is on the boundary of an element of Z. We suppose that if a disk of Z intersects $\mathrm{Bd}\,D$, the intersection is connected. We note that the inverses under f of the elements of Z give a null sequence Z' of mutually exclusive disks in $D-xy$.

Let a, b be points in different components of $\mathrm{Int}\,D-xy$ such that neither a nor b lies in an element of Z'. Let $(apb)'$ be an arc from a to b in $(D-xy)+\{p\}$. By replacing certain arcs in $(apb)'$ by arcs on the boundaries of elements of Z', one may obtain an arc apb in D that intersects Y only at p.

THEOREM 4.2. *Under the hypothesis of Theorem 4.1 there is a family of arcs from a to b such that no two of these arcs intersect except in their end points and for each point p of xy there is one of these arcs in $(D-Y)+\{p\}$.*

Proof. If $Y-xy$ were only a countable number of points it would be easy to get the prescribed family of arcs.

If g is a monotone map of D onto itself such that the elements of Z' (of Theorem 4.1) are the nondegenerate inverses under g of points of D, $g(Y-xy)$ is countable. We suppose g is the identity on xy. Let F be a family of arcs from $g(a)$ to $g(b)$ such that no two of these arcs intersect except in their end points and for each point p of xy there is an element of F in $(D-g(Y))+\{p\}$. The inverses of the elements of F are members of a family satisfying the conclusion of the theorem.

5. Piercing disks with tame arcs.

In this section we use the results of the preceding sections to show that each disk in E^3 is pierced by a tame arc.

THEOREM 5.1. *Suppose S is a 2-sphere, E is a tame disk, and X is a Sierpinski curve in $S\cdot E$. Then S can be pierced by a tame arc at each point of X.*

Proof. Let x be a given point of X. With no loss of generality we suppose that X lies in a horizontal plane P and xy is a straight line interval in X that does not intersect the boundary of a component of $P-X$ except possibly at x.

Let P' be a vertical plane containing xy and D be a circular disk in P'

with xy as diameter. We shall show that there is a tame arc in D that pierces S at x.

For each ϵ, $S - X$ has only a finite number of components of diameter greater than ϵ and no one of these components has any point of $xy - x$ as a limit point. Hence, for each $\epsilon > 0$ and each point p of $\operatorname{Int} xy$ there is a neighborhood N of p such that N intersects no component of $D \cdot S - xy$ of diameter greater than ϵ.

It follows from Theorem 4.1 that there are points a, b of D on opposite sides of xy such that for each point p of xy there is an arc apb in D that intersects S only at p. The arc apb pierces the plane P at p and it follows from Theorem 3.4 that if $p \neq x$, apb pierces S at p. Hence a and b belong to different components of $E^3 - S$. Hence, even in the case where $p = x$, apb pierces S at p.

Question. Theorem 4.2 can be used to show that there is a family of mutually exclusive tame arsc such that for each point p of xy there is a member of the family that pierces S at xy. One might wonder if there is such a family with xy replaced by X. A more general question—is there a family F of mutually exclusive tame arcs such that for each point of S at which S can be pierced by a tame arc, there is a member of F piercing S there?

THEOREM 5.2. *Each 2-sphere S in E^3 contains a 0 dimensional G_δ set Y such that S can be pierced by a tame arc at each point of $S - Y$.*

Proof. It follows from [4] that for each integer i, S contains a Sierpinski curve X_i such that X_i lies in a tame 2-sphere and each component of $S - X_i$ is of diameter less than $1/i$. It follows from Theorem 5.1 that S can be pierced by a tame arc at each point of the F_σ set $X_1 + X_2 + \cdots$. The set Y of Theorem 5.2 is $S - (X_1 + X_2 + \cdots)$.

The following theorem is a consequence of Theorem 5.2 and Theorem 2.1.

THEOREM 5.3. *Each disk D in E^3 contains a 0 dimensional G_δ set Y such that D can be pierced by a tame arc at each point of* $\operatorname{Int} D - Y$.

Questions. Suppose S is a 2-sphere and Y is the set of points of S at which S cannot be pierced by a tame arc. What can be said about the category of Y? Is it a 0 dimensional F_σ set?

If xy is a tame arc in a 2-sphere S, can S be pierced by a tame arc at each point of xy? (Note. Since this paper was written, David Gillman has answered this question in the affirmative.)

Suppose S is a 2-sphere in E^3, L is a straight line, and $\epsilon > 0$. Is there a homeomorphism h of E^3 onto itself such that h moves no point more than ϵ, h is the identity outside an ϵ neighborhood of L, and $H(S) - L$ is finite. An easy positive answer to this question would have simplified the approximation work in [2]. (Note. Since this paper was written, an affirmative answer to the question was announced in the 1962 Notices, Abstract 589-13. The methods used however were not easy.)

UNIVERSITY OF WISCONSIN.

REFERENCES.

[1] R. H. Bing, "A simple closed curve that pierces no disk," *Journal de Mathematiques Pures Appliquées*, vol. 35 (1956), pp. 337-343.

[2] ———, "Approximating surfaces with polyhedral ones," *Annals of Mathematics*, vol. 65 (1957), pp. 456-483.

[3] ———, "A surface is tame if its complement is 1-ULC," *Transactions of the American Mathematical Society*, vol. 101 (1961), pp. 284-305.

[4] ———, "Each disk in E^3 contains a tame arc," *American Journal of Mathematics*, vol. 84 (1962), pp. 583-590.

[5] Ralph H. Fox and Emil Artin, "Some wild cells and spheres in three-dimensional space," *Annals of Mathematics*, vol. 49 (1948), pp. 979-990.

[6] W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1948.

[7] R. L. Moore, "Concerning upper semicontinuous collections of continua," *Transactions of the American Mathematical Society*, vol. 27 (1925), pp. 416-428.

THE AUTOMORPHISMS OF THE GROUP OF SIMILITUDES AND SOME RELATED GROUPS.*[1]

By MARÍA J. WONENBURGER.[2]

TO THE MEMORY OF DN. JULIO REY PASTOR.

In [5] J. Dieudonné has determined the automorphisms of the orthogonal group, the group of rotations and their corresponding projective groups under the assumptions that the base field has characteristic not 2, the dimension of the vector space is sufficiently large and the index of the quadratic form is greater than zero. It has been proved that his results hold true without the restriction on the index of the quadratic form. For the case of the orthogonal group this was shown by C. Rickart in [8], for the projective group it was proved by J. Walter in [9] and for the case of the group of rotations and the projective group of rotations it was proved in [13]. Our purpose now is to determine the automorphisms of the group of similitudes, the group of proper similitudes and their corresponding projective groups. The general method consists in a reduction to Dieudonné's case, namely, it is shown first that, if an automorphism of any of the groups under consideration takes the subgroup of rotations or the cosets defined by the subgroup of rotations into themselves, the automorphisms of these groups are determined by Dieudonné's results (see Lemma 2 below and its corollaries). Hence our task is to prove the invariance of the group of rotations or the projective group of rotations under the automorphisms of the groups under consideration.

The case of the group of similitudes or the group of proper similitudes is settled first. If we deal with an odd dimensional vector space the projective group of similitudes is isomorphic to $O^+(Q)$, so there is nothing to prove. When the vector space is even dimensional the study of the automorphisms of the projective group of similitudes and the projective group of proper similitudes requires a more careful investigation of the new involutions appearing in these groups in order to characterize the cosets of the $(n-2, 2)$ involutions.

---

* Received May 22, 1962.

[1] Most of the results in this paper are generalizations of results included in the author's dissertation written under the direction of Professor N. Jacobson and presented to Yale University in 1957.

[2] Postdoctorate Fellow (of the National Research Council of Canada) at Queen's University.

Once these cosets have been characterized the results follow from the corollaries of Lemma 2. Since our characterization of the cosets of $(n-2,2)$ involutions excludes the case when K has 3 or 5 elements, we study first the automorphisms of the projective groups under consideration when the base field is a finite field.

1. Let M be a left vector space over a commutative field K and Q a non-degenerate quadratic form on M. *It is always assumed that the field K has characteristic not 2.* The elements of K will be always denoted by small greek letters, the elements of M by small latin letters and the semi-linear transformations of M by capital latin letters. The image of $x \in M$ under a transformation T will be denoted by xT, and, when the scalar multiplication by $\alpha \in K$ is considered as a linear transformation it is denoted by α_L.

A semi-linear transformation T of M relative to the automorphism σ of K is called a semi-similitude if for any $x \in M$

$$Q(xT) = \rho(Q)x))^{\sigma},$$

where $\rho \neq 0$ is a fixed element of K which depends on T and is called the ratio of T. In particular, if σ is the identity, T is called a similitude. When M is odd dimensional every similitude is the scalar multiple of a rotation. When M is even dimensional, say $\dim M = 2m$, the determinant of the matrix representing the similitude T of ratio ρ with respect to any basis is either ρ^m or $-\rho^m$. In the first case T is called a proper similitude and when the determinant is $-\rho^m$ T is an improper similitude.

The similitudes form a group, which we will denote by $S(Q)$; when the space M is even dimensional this group contains as a normal subgroup of index 2 the group of proper similitudes, which we denote by $S^+(Q)$. The centres of these groups consist of the scalar multiplications with the only exception of the case of the group of proper similitudes of a two dimensional space (see e.g. [3]).

The similitudes of ratio 1 form the orthogonal group $O(Q)$ and the group of proper similitudes of ratio 1 is the group of rotations $O^+(Q)$. The centre of $O(Q)$ consists of $\pm 1_L$ with the only exception of the case when $\dim M = 2$, K is the prime field with 3 elements and Q has index 1 (see e.g. [1, page 132]).

Let K' be the multiplicative group of the field K. Then the projective groups $PS(Q)$ and $PS^+(Q)$ are defined as

$$PS(Q) = S(Q)/K', \qquad PS^+(Q) = S^+(Q)/K',$$

which, with the exception of the case of the group $S^+(Q)$ of a two dimensional space, are the factor groups of $S(Q)$ and $S^+(Q)$ by their centres. Similarly, if Z is the center of $O(Q)$, $Z = \pm 1_L$, with the only exception of the 2 dimensional case mentioned above, and

$$PO(Q) = O(Q)/Z \text{ and } PO^+(Q) = O^+(Q)/Z \cap O^+(Q).$$

We are going to use the following known results:

(i) When $\dim M = n \geqq 4$ every automorphism ϕ of $O(Q)$ may be written in the form

$$\phi(G) = T^{-1}GT\chi(G) \text{ for every } G \in O(Q),$$

where $G \rightarrow \chi(G)$ is a representation of $O(Q)$ in the multiplicative group $\{1_L, -1_L\}$ and T is a semi-similitude of Q. (Cf. [5], [8] and [6]).

(ii) When $\dim M \geqq 5$ every automorphism of $O^+(Q)$ is the restriction of an automorphism of $O(Q)$. (See [5], [6] and [13]).

(iii) When $\dim M \geqq 4$ every automorphism of $PO(Q)$, the projective orthogonal group, is induced by an automorphism of $O(Q)$. (See [5], [9] and [6]).

(iv) When $\dim M \geqq 5$ but $\dim M \neq 8$ every automorphism of $PO^+(Q)$ is induced by an automorphism of $O^+(Q)$, (cf. [5], [6] and [13], and also Theorem 4 below).

(v) When $\dim M \geqq 6$ and $\neq 8$, and K is a finite field every automorphism of $P\Omega(Q)$ the projective commutator group is induced by an automorphism of $O^+(Q)$. (Cf. [5] or [6]).

Any element of order 2 of any of the groups under consideration is called an involution of that group. It is well-known that, if U is an orthogonal involution, that is, U is an orthogonal transformation and $U^2 = 1_L$, then the space M splits in a direct sum of two subspaces,

$$M = M^+ \oplus M^-,$$

such that U leaves invariant every element of M^+ and takes every vector of M^- into its negative. The subspaces M^+ and M^- are non-isotropic and mutually orthogonal. If $\dim M^+ = p$, then $\dim M^- = n - p$ and U is called a $(p, n-p)$ involution. It is clear that, if a linear transformation commutes with U, it must leave the spaces M^+ and M^- invariant and that there exist invertible linear transformations anticommuting with U if, and only if, $\dim M^+ = \dim M^-$ since such transformation must take M^+ into M^- and M^- into M^+. Now suppose that T is a similitude which commutes with U, the restriction T_1 (T_2) of T to M^+ (M^-) is a similitude of this subspace with respect to the restriction Q_{M^+} (Q_{M^-}) of Q to M^+ (M^-). Hence, such T will be represented by $T_1 \times T_2$, where T_1 and T_2 have the same ratio. Conversely, if $T_1 \in S(Q_{M^+})$ and $T_2 \in S(Q_{M^-})$ have the same ratio ρ, we can compose both of them to obtain a similitude $T_1 \times T_2 = T$ of $M = M^+ \oplus M^-$ by taking, for any $x = x_1 + x_2$, $x_1 \in M^+$, $x_2 \in M^-$, $Tx = T_1x_1 + T_2x_2$. Then T is a similitude of ratio ρ with respect to Q and it is proper if T_1 and T_2 are both proper or both improper, otherwise T is improper.

The coset of $PS(Q)$ defined by a similitude T will be denoted by $\bar{T}$. It is clear that $\bar{T}_1\bar{T}_2 = \bar{T}_2\bar{T}_1$ implies $T_1T_2 = T_2T_1$ or $T_1T_2 = -T_2T_1$.

2. Let $\Omega(Q)$ be the commutator group of the orthogonal group $O(Q)$. It is well-known that if $\dim M \geqq 3$, the centralizers of $\Omega(Q)$ in $\Omega(Q)$, $O^+(Q)$ and $O(Q)$ consist either of the identity transformation 1_L or of $\pm 1_L$ (see [1, Th. 3.23]). In what follows we will need to know the centralizers of $P\Omega(Q)$, $PO^+(Q)$ and $PO(Q)$ in $PS(Q)$ or $PS^+(Q)$. Now, if we establish that the centralizer of $P\Omega(Q)$ in $PS(Q)$ consist of the coset of the identity, it will follow that the same is true for the centralizers of $PO^+(Q)$ and $PO(Q)$. The essence of the proof is Artin's proof of the theorem just quoted.

LEMMA 1. *Let $\Omega(Q)$ be the commutator group of the orthogonal group $O(Q)$, Q a non-degenerated quadratic form over the vector space M of dimension greater than 2. Then the centralizer in $PS(Q)$ of the group of coset defined by $\Omega(Q)$ consists of the coset of the identity.*

Proof. We consider two different cases.

Case 1. The base field K contains more than 5 elements. In this case, given a 2 dimensional vector space over K with a non-degenerate quadratic form Q_1 the group $\Omega(Q_1)$ contains elements which are not involutions (see e. g. [13]). Now, if N is a non-isotropic two dimensional subspace of M with respect to Q, the commutator group $\Omega(Q)$ contains the group $\Omega(Q_N) \times 1_{N^\perp}$, where Q_N is the restriction of Q to N and $1_{N^\perp}$ is the identity transformation of the vector space $N^\perp$. Let G' be an element of $\Omega(Q_N)$ which is not an involution, then $G = G' \times 1_{N^\perp} \in \Omega(Q)$. If T is a similitude whose coset $\bar{T}$ belongs to the centralizer of $P\Omega(Q)$ in $PS(Q)$ $\bar{G}\bar{T} = \bar{T}\bar{G}$, and we must have $GT = -TG$ or $TG = GT$. We show first that $-TG = GT$ is impossible, for then, if $x \in N^\perp$,

$$xT = xGT = -xTG, \text{ that is, } (xT)G = -xT,$$

but no element of M is taken into its negative by G. Hence $TG = GT$, which implies that T takes $N^\perp$ into itself and, consequently, T takes N into itself. Since any one dimensional space of M can be expressed as the intersection of two non-isotropic subspaces of dimension 2, T leaves invariant all the one dimensional subspaces and, therefore, T is a scalar multiplication. In other words T belongs to the coset of the identity.

Case II. Q has index greater than zero (this is always the case if K is a finite field when $\dim M \geqq 3$). If x is any isotropic vector, let N be a non-isotropic 3 dimensional subspace which contains x and let $y \in N$ be a non-isotropic vector orthogonal to x, then there is a rotation G_α of N which leaves x invariant and takes y into $y + \alpha x$ (see [1, p. 133]). The rotation G_α, $\alpha \neq 0$, being the square of $G_{\alpha/2}$ belongs to the commutator group $\Omega(Q_N)$ and hence to $G_\alpha \times 1_{N^\perp} \in \Omega(Q)$. Now, if there were a vector $z \in N$ such that $xG_\alpha = -z$, then $zG_{2\alpha} = zG_\alpha{}^2 = z$; hence, $G_{2\alpha}$ being a rotation of a 3 dimensional space which leaves invariant the plane spanned by x and z, should be the identity mapping (see [1, Th. 3.17]) but this is impossible, because $yG_{2\alpha} = y + 2\alpha x$. So, if an element $T \in PS(Q)$ commutes with the coset of $T_\alpha = G_\alpha \times 1_{N^\perp}$, we must have $TT_\alpha = T_\alpha T$; therefore, T must leave invariant the subspace N_1 spanned by x and $N^\perp$, because N_1 is the subspace of vectors invariant under T_α. Since Kx is the subspace of N_1 orthogonal to N_1, T must take Kx into itself. Hence T leaves invariant every isotropic one dimensional subspace and, consequently, it is a scalar multiple of the identity (see [1, Th. 3.18]).

LEMMA 2. *Let M be a vector space with a non-degenerate quadratic form Q and* $\dim M > 4$. *If ϕ is an automorphism of the group $S(Q)$ (or $S^+(Q)$) which takes $O^+(Q)$ into itself, ϕ can be written in the form*

$$\phi(T) = V^{-1}TV\chi(T)$$

where $\chi(T)$ is a representation of $S(Q)$ ($S^+(Q)$) into the multiplicative group K' and V is a semi-similitude of Q.

Proof. By the result quoted in Section 1 as (ii) the restriction of ϕ to $O^+(Q)$ can be written in the form

$$\phi(G) = V^{-1}GV\chi(G)$$

Let σ be the automorphism of $S(Q)$ ($S^+(Q)$) defined by

$$\sigma(T) = VTV^{-1}.$$

Then the restriction of the automorphism $\sigma\phi$ to $O^+(Q)$ gives

$$\sigma\phi(G) = G\chi(G) \tag{1}$$

Let T be any similitude (proper similitude) and write $\sigma\phi(T) = TX_T$. Now, if G is any rotation so is $T^{-1}GT$, hence, by (1)

$$\sigma\phi(T^{-1}GT) = T^{-1}GT\chi(T^{-1}GT) \tag{2}$$

On the other hand,

$$\sigma\phi(T^{-1}GT) = \sigma\phi(T^{-1})\sigma\phi(G)\sigma\phi(T) = X_T^{-1}T^{-1}GTX_T\ \chi(G). \tag{3}$$

Comparing (2) and (3) we deduce that X_T commutes or anticommutes with the elements $T^{-1}GT$ for all $G \in O^+(Q)$; therefore, since $T^{-1}GT$ runs over $O^+(Q)$, the coset $\bar{X}_T$ of X_T in $PS(Q)$ $(PS^+(Q))$ belongs to the centralizer of $PO^+(Q)$, that is, X_T is a scalar multiplication which will be called $\chi(T)$. So $\sigma\phi(T) = T\chi(T)$ and $\phi(T) = V^{-1}TV\chi(T)$.

Corollary 1. *If* $\dim M \geqq 4$, *any automorphism* ϕ *of* $S(Q)$ *which takes* $O(Q)$ *into itself has the form described in the lemma.*

Proof. It is enough to substitute $O(Q)$ for $O^+(Q)$ in the proof of the lemma and to use (i) instead of (ii).

Corollary 2. *Let* $\dim M \geqq 4$. *Any automorphism* ϕ *of* $PS(Q)$ *which takes* $PO(Q)$ *into itself is induced by an automorphism of* $S(Q)$, *that is,*

$$\phi(\bar{T}) = \overline{V^{-1}TV}, \text{ where } V \text{ is a semi-similitude of } Q.$$

The proof is the same as for the lemma with the obvious modifications, namely, $\chi(T)$ does not appear and instead of (ii) we use (iii).

Corollary 3. *Let* $\dim M = 2m$, $m > 2$ *but* $m \neq 4$. *Any automorphism of* $PS(Q)$ *(or* $PS^+(Q)$*) which takes* $PO^+(Q)$ *into itself is induced by an automorphism of* $S(Q)$.

The proof uses (iv) instead of (iii) and by using (v) we obtain the next corollary.

Corollary 4. *Let* $\dim M = 2m \geqq 6$ *and* $\neq 8$. *If* K *is a finite field, every automorphism of* $PS(Q)$ *(or* $PS^+(Q)$*) which takes* $P\Omega(Q)$ *into itself is induced by an automorphism of* $S(Q)$.

3. We are ready now to find the automorphisms of $S(Q)$ and $S^+(Q)$.

THEOREM 1. *Let M be a vector space over the field K and $\dim M \geqq 4$ and let Q be a non-degenerate quadratic form on M. Every automorphism ϕ of the group of similitudes $S(Q)$ is of the form*

$$\phi(T) = V^{-1}TV\chi(T)$$

where V is a semi-similitude of Q and $\chi(T)$ is a representation of $S(Q)$ in K'.

Proof. On account of Corollary 1 of Lemma 2 we only need prove that every automorphism of $S(Q)$ takes $O(Q)$ into itself. Since the orthogonal group is generated by symmetries, that is, $(n-1,1)$ involutions, it is enough to show that every symmetry goes into an orthogonal transformation. If X is a symmetry, $(\phi(X))^2 = 1_L$; therefore, if $\phi(X)$ is not an orthogonal involution, $\phi(X)$ should be a similitude of ratio -1, so that $\phi(X)$ is a similitude of the type studied in [11]. In the proof of Proposition 4 of [11], it is shown that there exist orthogonal transformations anticommuting with the similitudes of such type. Now, if $Y\phi(X) = -\phi(X)Y$, it follows that

$$\phi^{-1}(Y)X = -X\phi^{-1}(Y)$$

which gives a contradiction, because, when $\dim M > 2$, there are no similitudes which anticommute with the symmetries.

THEOREM 2. *Let M be a vector space over the field K, $\dim M = 2m > 4$ and Q be a non-degenerate quadratic form over M. Every automorphism ϕ of the group $S^+(Q)$ is of the form*

$$\phi(T) = V^{-1}TV\chi(T),$$

where V is a semi-similitude and $\chi(T)$ is a representation of $S^+(Q)$ in K'.

Proof. It suffices to prove that under ϕ $O^+(Q)$ goes into itself. Since the group $O^+(Q)$ is generated by the $(n-2,2)$ involutions, if $O^+(Q)$ is not taken into itself by ϕ, at least one, say X, of the $(n-2,2)$ involutions must be taken by ϕ into an involution of $S^+(Q)$ which does not belong to $O^+(Q)$. As before we get that $\phi(X)$ has ratio -1 and is of the type studied in [11]. Now, if $\dim M = 4r$, we get a contradiction since, by [11, Proposition 4], there exist proper similitudes which anticommute with $\phi(X)$ whereas there are no similitudes anticommuting with the $(n-2,2)$ involutions when $\dim M > 4$. If $\dim M = 4r+2$, this argument does not apply because $S^+(Q)$ does not contain similitudes which anticommute with $\phi(X)$, but we know that in this case (see [11, Prop. 5]) the centralizer of the centralizer of $\phi(X)$

in $S^+(Q)$ consists of similitudes of the form $\alpha_L + \beta_L\phi(X)$ where $\alpha^2 - \beta^2 \neq 0$. If K has more than 3 elements, there exist elements $\alpha \neq 0$ and $\beta \neq 0$ such that $\alpha^2 - \beta^2 \neq 0$. On the other hand, when K has more than 3 elements, the centralizer of the centralizer of a $(n-2, 2)$ involution X consists of the elements α_L and $\alpha_L X$, where $\alpha \in K'$. This implies that the centralizer of the centralizer of X is mapped by ϕ into a proper subgroup of the centralizer of the centralizer of $\phi(X)$, which is a contradiction. Therefore $\phi(X) \in O^+(Q)$ and we can apply Corollary 3 of Lemma 2.

We have not taken care yet of the case when K has only 3 elements and $\dim M = 4r + 2$. Notice first that if there do not exist in $S^+(Q)$ involutions of ratio -1, any involution of $S^+(Q)$ is a rotation. On the other hand, if there are involutions of ratio -1, [10, Th. 5 and Corollary 3 of Prop. 4] imply that, when K is the prime field with 3 elements, the group $O^+(Q)$ is generated by the squares of proper similitudes. So that, in any case, any automorphism of $S^+(Q)$ must take $O^+(Q)$ into itself and we can use Corollary 3 of Lemma 2.

4. We assume in this section that K is a finite field and $\dim M > 4$. It is well-known that if K is a finite field and Q a quadratic form on a vector space M over K of dimension greater than 2, then Q has index greater than zero. It is also well-known that, when Q has index greater than zero and $\dim M \geqq 5$, $P\Omega(Q)$ is a simple group (cf. [4, Th. 2] and [6, p. 58]).

Let $(PS(Q))'$ and $(PS^+(Q))'$ be the first commutator groups of $PS(Q)$ and $PS^+(Q)$, respectively. Then

$$PO^+(Q) \supset (PS(Q))' \supset P\Omega(Q)$$

and

$$PO^+(Q) \supset (PS^+(Q))' \supset P\Omega(Q).$$

Now, because of our assumption, it follows that the second commutator groups of $PS(Q)$ and $PS^+(Q)$ are equal to $P\Omega(Q)$. Therefore any automorphism of $PS(Q)$ or $PS^+(Q)$ must take $P\Omega(Q)$ into itself and we can apply Corollary 4 of Lemma 2 to establish

THEOREM 3(I). *Let M be a vector space over a finite field K and $\dim M = 2m \geqq 6$, but $\dim H \neq 8$. Let Q be a non-degenerate quadratic form on M. Any automorphism of $PS(Q)$ or $PS^+(Q)$ is induced by an automorphism of $S(Q)$.*

5. In order to prove the assertion of Theorem 3(I) without the assumption that K is a finite field it will suffice to show that under any automorphism ϕ of $PS(Q)$ or $PS^{+}(Q)$ the cosets of $(n-2,2)$ involutions go into cosets defined by elements of the rotation group; for, this implies that $PO^{+}(Q)$ goes into itself and hence Corollary 3 of Lemma 2 can be applied. It will be shown in this section that, in fact, when K has more than 5 elements and $\dim M \geqq 5$ but $\neq 8$, any automorphism of $PS(Q)$ or $PS^{+}(Q)$ takes the cosets of $(n-2,2)$ involutions into cosets of $(n-2,2)$ involutions.

It will be convenient to describe the different kinds of involutions in $PS(Q)$. Let $T \in S(Q)$, then $\bar{T}$ is an involution in $PS(Q)$ if and only if $T^2 = \alpha_L$, such similitudes will be called projective involutions. The projective involutions of $PS(Q)$ can be divided in 3 classes:

1) T is a scalar multiple of an orthogonal involution, that is, $T = U\beta_L$, where U is an orthogonal involution. Then T has ratio β^2 and $\alpha = \beta^2$. We will say that $\bar{T}$ is the coset of an orthogonal involution since it contains two such involutions, namely, U and $-U$. If either U or $-U$ is a $(n-2,2)$ orthogonal involution we call $\bar{T}$ a $(n-2,2)$ coset.

2) T is a similitude of ratio ρ and $T^2 = -\rho_L$. Such similitudes will be called P-involutions. It was shown in [11, Lemma 1] that the P-involutions are always proper. When $\dim M = 4r+2$, the similitudes anticommuting with a P-involution are improper (see [11, Prop. 4]), so that, in particular, two P-involutions cannot anticommute.

3) T is a similitude of ratio ρ, $T^2 = \rho_L$, but ρ is not a square in K. When $\dim M \geqq 6$ and K has more than 3 elements, no automorphism of $PS(Q)$ or $PS^{+}(Q)$ takes a $(n-2,2)$ coset into the coset of such a T (see [12, Th. 5]; there is an omission in the statements of this theorem and of Lemma 9, for the lemma is only true if K has more than 3 elements and, therefore Theorem 5 only applies under this assumption).

In the rest of this section it is assumed that K contains more than 5 elements.

First of all we are going to characterize the $(n-2,2)$ cosets among the cosets of orthogonal involutions and state some properties of the P-involutions.

LEMMA 3. *Let M be a vector space with a non-degenerate quadratic form Q and* $\dim M > 4$. Let *U be an orthogonal involution of $S^{+}(Q)$ or $S(Q)$. If $(C_{PS^{+}}(\bar{U}))'$ $((C_{PS}(\bar{U}))')$ is the commutator of the centralizer of $\bar{U}$ in $PS^{+}(Q)$ $(PS(Q))$, then the center of $(C_{PS^{+}}(\bar{U}))'$ $((C_{PS}(\bar{U}))')$ contains elements which are not involutions if and only if $\bar{U}$ is a $(n-2,2)$ coset.*

Proof. Let Q^+ and Q^- be the restrictions of Q to the plus and minus spaces of the involution U. Then the centralizer of $\bar{U}$ in $PS^+(Q)$ $(PS(Q))$ contains the cosets of the transformations $T_1 \times T_2$, where T_1 and T_2 are similitudes of Q^+ and Q^-, respectively, with the same ratio and both proper or both improper (without restriction on the parity of T_1 and T_2). When the plus and minus spaces of U have different dimension the centralizer of $\bar{U}$ consists of such cosets only. Therefore, since $\dim M > 4$, when $\bar{U}$ is a $(n-2,2)$ coset the commutator of the centralizer of $\bar{U}$ is contained in the group of cosets determined by $O^+(Q^+) \times O^+(Q^-)$ and contains the cosets defined by $\Omega(Q^+) \times \Omega(Q^-)$. Hence, if U is a $(n-2,2)$ orthogonal involution, the center of $(C_{PS^+}(\bar{U}))'$ $((C_{PS}(\bar{U}))')$ contains an element which is not an involution, namely, $\overline{1_L \times G}$, where $G \in \Omega(Q^-)$ and $G^2 \neq 1_L$.

On the other hand if U is an $(n-p,p)$ orthogonal involution and $n-p \neq 2$, $p \neq 2$, the commutator of the centralizer of $\bar{U}$ in $PS^+(Q)$ or $PS(Q)$ is always contained in the group of cosets defined by $O(Q^+) \times O(Q^-)$ and contains the cosets defined by $\Omega(Q^+) \times \Omega(Q^-)$. Hence the center of $(C_{PS^+}(\bar{U}))'$ or $(C_{PS}(\bar{U}))'$ consists of the coset of the identity or of such coset and the coset defined by $(-1_L) \times 1_L$.

From this lemma and [12, Th. 5], quoted above, we get the following corollary.

Corollary. *If* $\dim M > 4$, *under an automorphism of* $PS(Q)$ or $PS^+(Q)$ *a* $(n-2,2)$ *coset goes into a* $(n-2,2)$ *coset or into the coset of a P-involution.*

Lemma 4. *Let T be a P-involution. Then any two $(n-2,2)$ orthogonal involutions commuting with T and commuting with each other have their minus spaces orthogonal to each other.*

Proof. It is well-known that two different $(n-2,2)$ orthogonal involutions which commute must have their minus spaces orthogonal to each other or the intersection of these minus spaces is a non-isotropic one dimensional space. If we were in the latter case, since both involutions commute with T, T would leave invariant this one dimensional space. This is impossible, because, if ρ is the ratio of T, $(x,xT) = \rho^{-1}(xT,xT^2) = -(xT,x)$ implies $(x,xT) = 0$, that is, T takes a non-isotropic vector into another vector orthogonal to it.

Corollary. *Let* $\dim M = 2m > 4$. *Let* $\{U_i\}$ *be a set of* $(n-2,2)$ *orthogonal involutions which commute with T and with each other, then there are at most m elements in the set* $\{U_i\}$.

We are going to assume that $\dim M = 2m > 4$ and that there exists an automorphism ϕ of $PS(Q)$ or $PS^+(Q)$ which takes a $(n-2,2)$ coset $\bar{U}$ into the coset of a P-involution T. Let U be the $(n-2,2)$ orthogonal involution of $\bar{U}$, let x_1, x_2 be an orthogonal basis of its minus space and $x_3, x_4, \cdots, x_{2m}$ an orthogonal basis of its plus space. Let us define U_i, $i = 2, 3, \cdots, 2m$, as the $(2m-2,2)$ orthogonal involution whose minus space is spanned by the vectors x_1 and x_i. Then $U = U_2$, the $2m-1$ involutions U_i commute with each other and have the property that the product of any two is also a $(n-2,2)$ orthogonal involution. Now, by the Corollary of Lemma 3, we know that the cosets $\phi(U_i)$, $i = 2, \cdots, 2m$, are $(n-2,2)$ cosets or cosets of P-involutions and the product $\phi(U_i)\phi(U_j)$, $i \neq j$, is also a $(n-2,2)$ coset or the coset of a P-involution.

LEMMA 5. *Let U_i, $i = 2, 3, \cdots, 2m$, $2m > 4$, be the set of $(n-2,2)$ involutions described above and let $\phi(U_i) = \bar{T}_i$. Then*

(1) *either all the T_i, $i = 2, \cdots, 2m$ commute with each other, or*

(2) *any two different T_i anticommute.*

Proof. We can assume that if T_i is not a P-involution, it is a $(n-2,2)$ orthogonal involution. It is clear that if $\dim M = 4r + 2$ all the T_i commute with each other (see the properties of the P-involution given at the beginning of this section).

On the other hand, if $\dim M = 4r$, it seems possible that some pairs T_i, T_j commute and some others anticommute; however, we will show that if two different T_i commute all of them must commute. To show this we prove first that if T_i and T_j commute, then one of the cosets $\bar{T}_i$, $\bar{T}_j$ or $\bar{T}_i\bar{T}_j$ is a $(4r-2,2)$ coset. We only need prove that, if neither T_i nor T_j is a $(4r-2,2)$ orthogonal involution, $\overline{T_iT_j}$ is a $(4r-2,2)$ coset. Let ρ_i and ρ_j be the ratios of T_i and T_j, then T_iT_j has ratio $\rho_i\rho_j$ and $(T_iT_j)^2 = \rho_i\rho_j$; hence, T_iT_j is not a P-involution and, consequently, $\overline{T_iT_j}$ must be a $(4r-2,2)$ coset. Now, since every $\bar{T}_k$ commutes with $\bar{T}_i$, $\bar{T}_j$ and $\bar{T}_i\bar{T}_j$, T_k must leave invariant the $4r-2$ and 2 dimensional spaces defined by the $(4r-2,2)$ coset; hence the T_k which are P-involutions induce P-involutions in these spaces and we know that in a space of dimension $4r-2$ two P-involutions cannot anticommute. This proves the lemma and, in particular, if all the T_k anticommute with each other there cannot be $(n-2,2)$ involutions among them.

LEMMA 6. *Let $\dim M = 2m$, $m > 2$ but $m \neq 4$. Then any automorphism ϕ of $PS(Q)$ or $PS^+(Q)$ takes a $(2m-2,2)$ coset into a $(2m-2,2)$ coset.*

Proof. Assume that U is a $(2m-2,2)$ orthogonal involution such that $\phi(\bar{U}) = \bar{T}$, where T is a P-involution. Let the U_i, $i = 2, \cdots, 2m$, be defined as before. We are going to show that, when $m > 2$, alternative (1) leads to contradiction and alternative (2) can only occur when $\dim M = 8$.

Case 1. Let $\bar{T}_i = \phi(\bar{U}_i)$ and assume that all the T_i commute with each other. We have seen that this implies that if $\bar{T}_j$ is not a $(n-2,2)$ coset then $\bar{T}_2\bar{T}_j$ is a $(n-2,2)$ coset. By Lemma 4, at most one of the T_j, say T_3, is an orthogonal involution, consequently $\bar{T}_2\bar{T}_j$, $j = 4, 5, \cdots, 2m$, is a $(2m-2,2)$ coset, so that $T_j = T_2 V_j \alpha_{jL}$, where V_j is a $(2m-2,2)$ orthogonal involution which commutes with T_2. As for T_3 either $T_3 = T_2 V_3 \alpha_{3L}$ or $T_3 = V_3$. Now, the V_j, $j = 3, \cdots, 2m$, commute with each other and with T_2, therefore, by the Corollary of Lemma 4, there are at most m different V_j. Hence $m \geqq 2m-2$, that is, $m \leqq 2$.

Case 2. Any two different T_i anticommute. But, if $\dim M = 2^k p$, p not divisible by 2, the maximal number of anticommuting non-singular linear transformations is $2k+1$ (see e.g. [7, Th. 2]; since in our case $T_i^2 = \alpha_L$ the result can be proved by using the theory of Clifford algebras as in Section 6 below). Since the number of T_i is $2^k p - 1$ we must have $2^k p - 1 \leqq 2k + 1$ and if $2^k p > 4$, this is only possible when $k = 3$ and $p = 1$, that is to say, if $\dim M = 8$.

From Lemma 6 and Corollary 3 of Lemma 2 we get

THEOREM 3(II). *Let Q be a non-degenerate quadratic form over a vector space M and $\dim M = 2m > 4$ but $\dim M \neq 8$. If the base field K has more than 5 elements, every automorphism ϕ of $PS(Q)$ or $PS^+(Q)$ is of the form*

$$\phi(\bar{T}) = \overline{VTV^{-1}},$$

where V is a semi-similitude of Q.

6. In this section we are going to study the case $\dim M = 8$ which is left out in Thseorem 3(I) and Theorem 3(II). When we deal with $PS^+(Q)$ there can actually exist automorphisms different from the ones described in these theorems (cf. [2]). We will show first that there are not exceptional automorphism of $PO^+(Q)$ if K is a finite field and then it will be proved that even when there exist exceptional automorphisms of $PS^+(Q)$ they cannot be extended to $PS(Q)$. We recall that in the determination of the automorphisms of $PO^+(Q)$ (cf. [4]), when $\dim M = 8$, the trouble arises, as for $PS^+(Q)$, from the fact that it is possible that a $(n-2,2)$ coset is taken by an automorphism ϕ into the coset of a P-involution (the orthogonal P-involutions are the involutions of the second kind of Dieudonné). Now, if it can be proved that the $(n-2,2)$ cosets of $PO^+(Q)$ go into $(n-2,2)$ cosets, the usual proof can be carried through showing that the automorphisms of $PO^+(Q)$ are induced by automorphisms of $O^+(Q)$.

THEOREM 4. *Let M be a 8-dimensional vector space over K and Q a non-degenerate quadratic form over K. Then, if K is a finite field, every automorphism of $PO^+(Q)$ is induced by an automorphism of $O^+(Q)$.*

Proof. When K is a finite field it contains only two different quadratic classes and any subspace of M of dimension greater than 2 contains isotropic vectors; consequently, if U is a $(n-2,2)$ orthogonal involution such that the restriction Q^- of Q to the minus space of U has a square discriminant, U can be expressed as the product of two $(n-2,2)$ orthogonal involutions V_1 and V_2 which commute with each other and such that the discriminants of the restrictions of Q to the minus spaces of V_1 and V_2 are not squares.

To establish our theorem we only need prove that under any automorphism ϕ of $PO^+(Q)$ a $(n-2,2)$ coset $\bar{U}$ cannot be taken into the coset of a P-involution; for then $\bar{U}$ must be taken into another $(n-2,2)$ coset (see [4, Section 39]). Hence, if there are no orthogonal P-involutions there is nothing to prove. If there exist orthogonal P-involutions, Q has a square discriminant and this implies that the coset of a $(n-2,2)$ involution U belongs to $P\Omega(Q)$, the commutator group of $PO^+(Q)$, if and only if the restriction of Q to the minus space of U has a square discriminant, because the spin-norms of U and $-U$ are equal to the discriminant of Q^- (see e.g. [1, Th. 5.17]).

Now, if T is an orthogonal P-involution, $\bar{T} \in P\Omega(Q)$. (This can be seen by computing the spin-norm of T or directly, for, since there exists an orthonormal basis x_i, y_i, $i=1,2,3,4$ of Q such that $x_iT=y_i$ and $y_iT=-x_i$ (cf. e.g. [11, Prop. 1]) $T=T_1ST_1^{-1}S^{-1}$, where T_1 and S are defined as follows,

$$\begin{array}{lll} x_jT_1=y_j, & y_jT_1=-x_j, & \text{for } j=1,2, \text{ and} \\ x_jT_1=x_j, & y_jT_1=y_j, & \text{for } j=3,4, \text{ and} \\ x_iS=y_{i+2}, & y_iS=x_{i+2}, & \text{where the indexes } i+2 \end{array}$$

should be computed module 4). So, if the restriction of Q to the minus space of the $(n-2,2)$ involution U does not have a square discriminant, then $\phi(\bar{U})$ must be a $(n-2,2)$ coset. Now, if this restriction has a square discriminant, $U=V_1V_2$, where $V_1, V_2 \notin P\Omega(Q)$ and $V_1V_2=V_2V_1$; hence $\phi(\bar{U})=\phi(\bar{V}_1)\phi(\bar{V}_2)$, that is, $\phi(\bar{U})$ is the product of two commuting $(n-2,2)$ cosets and, consequently, it cannot be a P-involution.

Remark. [2, Corollary 2 to Th. 1 and Th. 8] imply that under the conditions of Theorem 4 above, when Q has a square discriminant, $PS^+(Q)$ does have exceptional automorphisms.

We return now to the study of $PS(Q)$ and consider two cases.

Case 1. *K is a finite field.* In this case it is easily seen that any 2-dimensional space over K contains similitudes of any ratio with respect to any quadratic form, hence the same is true of M with respect to Q. Then [10, Cor. 3 of Prop. 4] implies that the first commutator group of $PS(Q)$ is $PO^+(Q)$, and, therefore, every automorphism ϕ of $PS(Q)$ induces an automorphism in $PO^+(Q)$. Now, Theorem 4 shows that the proof of Lemma 2 can be applied to the present case to obtain $\phi(\bar T) = \overline{V^{-1}TV}$.

Case 2. *K has more than 5 elements.* Let us suppose that under an automorphism ϕ of $PS(Q)$ a $(n-2,2)$ coset $\bar U$ is not taken into a $(n-2,2)$ coset. Then, by the corollary of Lemma 3, $\phi(\bar U) = \bar T$, where T is a P-involution. Consider the set of $(n-2,2)$ orthogonal involutions U_i, $i = 2, 3, \cdots, 8$ defined in the preceding section; the proof of Lemma 6 shows that $\phi(\bar U_i) = \bar T_i$, where the T_i are P-involutions and any two of them anticommute. Let V be the symmetry whose minus space is spanned by the vector x_1, then $U_i V = V U_i$ and, consequently, if $\phi(\bar V) = \bar W$, either $WT_i = T_i W$ or $WT_i = -T_i W$, in other words the linear transformation W commutes with some of the T_i and anticommute with the others. Let A_j, $j = 2, 3, \cdots, 8$ be 8×8 matrices representing the linear transformations T_j with respect to a certain basis of M, then $A_j^2 = -\rho_j I$ where I is the 8×8 identity matrix. This means that, if N is a 6-dimensional vector space over K and Q' is a quadratic form on N such that there exists an orthogonal basis v_i, $i = 1, 2, \cdots, 6$, with $Q'(v_i) = -\rho_{i+1}$, there exists a homomorphism σ of the Clifford algebra $C(Q')$ defined by Q' into the enveloping algebra of the matrices A_j, $j = 2, 3, \cdots, 7$, namely, $\sigma(v_i) = A_{i+1}$, $i = 1, 2, \cdots, 6$. Since the Clifford algebra $C(Q')$ is simple and $\dim C(Q') = 2^6 = 8 \cdot 8$, σ is an isomorphism of $C(Q')$ onto the algebra of 8×8 matrices with entries in K. Now, the only elements of $C(Q')$ which commute with some of the v_i and anticommute with the others are the elements of the Clifford group of the form $\alpha v_{i_1} v_{i_2} \cdots v_{i_k}$; hence, the only linear transformations which commute with some of the T_i and anticommute with the others are of the form $\alpha_L T_{i_1} T_{i_2} \cdots T_{i_k}$. But, it is obvious that $\phi(\bar V)$ cannot be $\bar T_{i_1} \bar T_{i_2} \cdots \bar T_{i_k}$, because $\bar V \neq \bar U_{i_1} \bar U_{i_2} \cdots \bar U_{i_k}$. Therefore any automorphism ϕ of $PS(Q)$ takes the $(6,2)$ cosets into $(6,2)$ cosets and must be of the form $\phi(\bar T) = \overline{V^{-1}TV}$.

We sum up these results together with the results of Theorems 3(I) and (II) in our last theorem.

THEOREM 5. *Let M be a vector space with a non-degenerate quadratic form Q and* $\dim M = 2r$, $r > 2$. *Every automorphism of $PS(Q)$, or $PS^{+}(Q)$ if* $\dim M \neq 8$, *is induced by an automorphism of $S(Q)$. When* $\dim M = 8$, *but there do not exist P-involutions, any automorphism of $PS^{+}(Q)$ is induced by an automorphism of $S^{+}(Q)$.*

QUEEN'S UNIVERSITY, KINGSTON, ONTARIO.

REFERENCES.

[1] E. Artin, *Geometric Algebra* (New York: Interscience Publishers, 1957).

[2] F. van der Blij and T. A. Springer, "Octaves and Triality," *Niew Archief voor Wiskunde* (3) 1960, pp. 158-169.

[3] N. Bourbaki, *Algèbre*, Chapter 9 (Paris: Hermann, 1959).

[4] J. Dieudonné, *Sur les Groupes Classiques* (Paris: Hermann, 1948).

[5] ———, "On the Automorphisms of the Classical Groups," *Memoirs of the American Mathematical Society*, No. 2 (1951).

[6] ———, *La Géométrie des Groupes Classiques* (Berlin: Springer-Verlag, 1955).

[7] H. Kestelman, "Anticommuting Linear Transformations," *Canadian Journal of Mathematics*, vol. 13 (1961), pp. 614-24.

[8] C. E. Rickart, "Isomorphic groups of linear transformations," *American Journal of Mathematics*, vol. 72 (1950), pp. 451-464; II, *American Journal of Mathematics*, vol. 73 (1951), pp. 697-716.

[9] J. Walter, "Isomorphisms between projective unitary groups," *American Journal of Mathematics*, vol. 77 (1955), pp. 805-844.

[10] M. J. Wonenburger, "The Clifford algebras and the group of similitudes," *Canadian Journal of Mathematics*, vol. 14 (1962), pp. 45-59.

[11] ———, "Study of certain similitudes," *Canadian Journal of Mathematics*, vol. 14 (1962), pp. 60-68.

[12] ———, "Study of a semi-involutive similitude," *Revista Matemática Hispano-Americana*, XX (1960), pp. 34-45.

[13] ———, "The automorphisms of the group of rotations and its projective group corresponding to quadratic forms of any index," *Canadian Journal of Mathematics* (1963).

ON ABELIAN INTEGRALS OF THE SECOND KIND AND MODULAR FUNCTIONS.*[1]

By Marvin Isadore Knopp.

I. Introduction. Let $\Gamma(j)$, $j = 2, 3, \cdots$, denote the principal congruence subgroup of level j of the modular group, $\Gamma(1)$. In the previous papers [2, 3] the author has constructed modular functions for $\Gamma(j)$ by first constructing abelian integrals of the second kind for $\Gamma(j)$ and then taking suitable linear combinations of these.

By an *abelian integral* for $\Gamma(j)$ we will always mean here a function $f(\tau)$ regular in the upper half-plane, $\operatorname{Im}(\tau) > 0$, and such that for $\operatorname{Im}(\tau) > 0$,

$$f(V\tau) = f(\tau) + C(V), \text{ for } V \in \Gamma(j),$$

where $C(V)$ is independent of τ. If in addition $f(\tau)$ has at most a polar singularity in the local uniformizing variable (cf. § III. 2) at each parabolic cusp of $\mathfrak{F}_j$, the fundamental domain of $\Gamma(j)$, then $f(\tau)$ is said to be of the *second kind.* If $f(\tau)$ is regular in the local uniformizing variable at each cusp, then $f(\tau)$ is said to be of the *first kind.* If $f(\tau)$ is an abelian integral of the second kind such that $C(V) = 0$ for all $V \in \Gamma(j)$, then $f(\tau)$ is said to be a *modular function* for $\Gamma(j)$.

In this paper we generalize the construction of [2, 3] to obtain a wider class of abelian integrals. The generalization is in two directions.

(1) While in [2, 3] we constructed abelian integrals of the second kind with arbitrary principal parts at *some* of the parabolic cusps of $\mathfrak{F}_j$, we here are able to obtain such integrals with arbitrary principal part at *all* of the cusps. This has been motivated in part by the work of Smart [14].

(2) By making use of the notion of the "supplementary series" (cf. [5]) we succeed in constructing abelian integrals of the *first kind* for $\Gamma(j)$ and in fact we obtain a basis for the finite dimensional vector space of abelian integrals of the first kind.

Combining the results cited above in (1) and (2) we find that we have

---

* Received May 24, 1962.

[1] Research supported in part by National Science Foundation grant number G-14362.

a *linear basis for the complex vector space of abelian integrals of the second kind for* $\Gamma(j)$ (Theorem 4.12)).

Further considerations show that if an abelian integral of the second kind is in particular a modular function for $\Gamma(j)$, then its expression in terms of the basis mentioned above is of a special kind (Theorem (5.1)), a kind in fact that gives rise to an explicit expression (5.5) of its Fourier coefficients in terms of the principal parts at the cusps of $\mathfrak{F}_j$ (§ V.1). This type of result is quite well known for automorphic forms of positive dimension (cf. [6, 7, 11, 16]) and in a few special cases for automorphic functions (cf. [9, 12, 13]). However, these results have been derived by means of various refinements of the circle method, while here the approach is quite different. Also, to the best of my knowledge the result has not previously been obtained for modular functions connected with the class of groups $\Gamma(j)$, $j \geqq 2$.

In § V.2 we construct modular functions for $\Gamma(j)$ each of which has at single pole of order at most $g+1$ at one of the parabolic cusps (here g is the genus of $\mathfrak{F}_j$).

II. The functions $\boldsymbol{\lambda_\nu(j, M; \tau)}$. Let ν be a positive integer, let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1)$, and let $\delta_\nu(M)$ be defined by

$$(2.1)\qquad \begin{aligned} \delta_\nu(M) &= (\tfrac{1}{10})\exp(-2\pi i\nu b/j), && \text{if } c \equiv 0,\ a \equiv d \equiv 1 \pmod j \\ &= (\tfrac{1}{10})\exp(2\pi i\nu b/j), && \text{if } c \equiv 0,\ a \equiv d \equiv -1 \pmod j \\ &= 0, \text{ otherwise.} \end{aligned}$$

Note that $\delta_\nu(-M) = \delta_\nu(M)$. We put

$$(2.2)\qquad a_m(\nu, j, M) = (\pi/8j) \sum_{k=1}^{\infty}{}^{M} k^{-1} A_{k,\nu}(m, M)\,(\nu/m)^{\frac{1}{2}} I_1(4\pi\sqrt{m\nu}/jk),$$

where

$$(2.3)\qquad A_{k,\nu}(m, M) = \sum_{\substack{0 \leqq h < jk \\ (h,k)=1}}^{M} \exp\left[\frac{-2\pi i}{jk}\,(\nu h' + mh)\right]$$

and I_1 is the modified Bessel function of the first kind, which can be given by

$$I_1(x) = \sum_{p=0}^{\infty} \frac{(x/2)^{2p+1}}{p!(p+1)!}$$

In (2.3) h' is an integral solution of the congruence $hh' \equiv -1 \pmod k$;

we put $k' = -k^{-1}(hh'+1)$. Then the superscript M appearing in (2.2) and (2.3) indicates that the sums are restricted by the condition

$$(2.4)\qquad \begin{pmatrix} h' & k' \\ k & -h \end{pmatrix} \equiv \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \pmod{j}$$

where the congruence is elementwise. The condition (2.4) is to be understood in the sense that each appropriate lower row $(k, -h)$ is to occur *exactly once* and that we are then to find *any* (h', k') such that $-hh' - kk' = 1$ and such that the matrix $\begin{pmatrix} h' & k' \\ k & -h \end{pmatrix}$ satisfies (2.4).

This ambiguity in the choice of h' and k' leads to the possibility that the sum $A_{k,\nu}(m, M)$ is not well defined; that is, it might depend upon the choice of h'. However h' is determined modulo jk and therefore $A_{k,\nu}(m, M)$ *is* well defined. For (h', k') can be replaced only by $(h' + tk, k' - th)$, with t an integer. But $h' + tk \equiv h'$, $k' - th \equiv k' \pmod{j}$, so that we can conclude that $tk \equiv th \equiv 0 \pmod{j}$. From this it follows that $t = t(-hh' - kk') \equiv 0 \pmod{j}$. Hence $h' + tk \equiv h' \pmod{jk}$.

The basic functions $\lambda_\nu(j, M; \tau)$ are now defined by

$$(2.5)\qquad \lambda_\nu(j, m; \tau) = \delta_\nu(M) e^{-2\pi i\nu\tau/j} + \sum_{m=1}^{\infty} a_m(\nu, j, M) e^{2\pi i m\tau/j}.$$

By the way in which they are defined these functions depend only upon the resideue class of M modulo $\Gamma(j)$. The functions λ_ν are regular in the upper half-plane, $\operatorname{Im}(\tau) > 0$. In order to see this it is necessary to have a nontrivial asymptotic estimate for the exponential sums $A_{k,\nu}(m, M)$ and in fact it turns out that as $k \to +\infty$

$$(2.6)\qquad A_{k,\nu}(m, M) = O(k^{\frac{1}{2}+\epsilon}), \quad \epsilon > 0.$$

It is apparent that the $A_{k,\nu}(m, M)$ are closely related to the classical Kloosterman sums and as one might expect (2.6) can be derived from the corresponding estimate for the Kloosterman sums (cf. [14, pp. 76-82]).

The functions that are treated in [3] form a special subclass of those we have defined in (2.4), that subclass, in fact, described by the additional restriction $b \equiv 0 \pmod{j}$. In order to see this, we let $l = jk$ in (2.3), note that $(d, j) = 1$ as a consequence of $b \equiv 0 \pmod{j}$, and use the fact that $h \equiv \pm d \pmod{j}$ and $(d, j) = 1$ together imply that $(h, j) = 1$. Hence in (2.3) $(h, jk) = (h, l) = 1$ and $hh' = -1 - kk' \equiv -1 \pmod{l}$. Thus h' is determined modulo l and $A_{k,\nu}(m, M)$ becomes

$$\sum_{\substack{0 \leqq h < l \\ (h, l) = 1 \\ h \equiv d \pmod{j}}} \exp\left[\frac{-2\pi i}{l}(\nu h' + mh)\right], \quad hh' \equiv -1 \pmod{l}.$$

A simple calculation now shows that the expression (2.2) for the coefficients a_m reduces in the case $b \equiv 0 \pmod{j}$ to the expression (1.2) of [3].

III. Abelian integrals of the second kind. 1. The most important feature of the functions $\lambda_\nu(j, M; \tau)$ is their behavior under transformation by elements of $\Gamma(1)$. This behavior is described in the following theorem.

THEOREM (3.1). (i) *Let* $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma(1)$. *Then for* $\operatorname{Im}(\tau) > 0$, $\lambda_\nu(j, M; \tau)$ *satisfies*

$$\lambda_\nu(j, M; V\tau) = \lambda_\nu(j, MV; \tau) + \omega_\nu(j, M; V), \tag{3.2}$$

where $\omega_\nu(j, M; V)$ *is independent of* τ.

(ii) *Furthermore if* $\gamma = 0$, *then* $\omega_\nu(j, M; V) = 0$, *and if* $\gamma \neq 0$, *then*

$$\omega_\nu(j, M; V) = \rho_\nu(j, M; V) + \sigma_\nu(j, M; V),^2$$

where

$$\begin{aligned} \rho_\nu(j, M; V) = {} & \frac{\pi^2 \nu}{8j^2} \sum_{k=1}^{\infty}{}^{MV} k^{-2} \sum_{\substack{0 \leqq h < jk \\ (h,k)=1}}{}^{MV} \exp(-2\pi i \nu h'/jk) \\ & - \frac{\pi^2 \nu}{8j^2} \sum_{k=1}^{\infty}{}^{M} k^{-2} \sum_{\substack{0 \leqq h < jk \\ (h,k)=1}}{}^{M} \exp(-2\pi i \nu h'/jk), \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} \sigma_\nu(j, M; V) = {} & \tfrac{1}{16} \lim_{K \to \infty} \Big\{ \sum_{k=1}^{[tK]}{}^{MV} \sum_{\substack{|m| \leqq K \\ (k,m)=1}}{}^{MV} \exp(-2\pi i \nu m'/k) \\ & - \sum_{\substack{(k,m) \in \mathcal{T}_V(K) \\ (k,m)=1}}{}^{M} \sum{}^{M} \exp(-2\pi i \nu m'/k) \Big\}, \end{aligned} \tag{3.4}$$

where t is a certain positive constant and $\mathcal{T}_V(K)$ a certain trapezoid in the $k-m$ plane (cf. [3, p. 663], for details).

Although this theorem is the basic one of the paper and although its proof is by no means trivial, we nevertheless omit the details since they are to a great extent a repetition of calculations to be found in several previous papers (cf. [2, 4, 8, 14] and especially [10] in which the general method first appeared). Some discussion of the case $\gamma = 0$ is in order, however, since the method here is different, although quite simple.

If $\gamma = 0$, then $V = \pm \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$, or $V\tau = \tau + \beta$, and

$$\pm MV = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & a\beta + b \\ c & c\beta + d \end{pmatrix}. \tag{3.5}$$

---

[2] In [3], p. 665, formula (2.7) the term corresponding to ρ_ν was mistakenly omitted.

Proceeding directly from (2.5) we have

$$(3.6)\qquad \begin{aligned}\lambda_\nu(j,M;V\tau) &= \delta_\nu(M)\,e^{-2\pi i\nu(\tau+\beta)/j}+\sum_{m=1}^{\infty}a_m(\nu,j,M)\,e^{2\pi i m(\tau+\beta)/j}\\ &= \delta_\nu(M)\,e^{-2\pi i\nu\beta/j}e^{-2\pi i\nu\tau/j}+\sum_{m=1}^{\infty}e^{2\pi i m\beta/j}a_m(\nu,j,M)\,e^{2\pi i m\tau/j}.\end{aligned}$$

By (2.1),

$$\begin{aligned}\delta_\nu(M)\,e^{-2\pi i\nu\beta/j} &= \tfrac{1}{16}e^{-2\pi i\nu(b+\beta)/j}, \text{ if } c\equiv 0,\ a\equiv d\equiv 1 \pmod{j}\\ &= \tfrac{1}{16}e^{-2\pi i\nu(-b+\beta)/j}, \text{ if } c\equiv 0,\ a\equiv d\equiv -1 \pmod{j}\\ &= 0, \text{ otherwise,}\end{aligned}$$

and, on the other hand,

$$\begin{aligned}\delta_\nu(\pm MV) &= \tfrac{1}{16}e^{-2\pi i\nu(a\beta+b)/j} = \tfrac{1}{16}e^{-2\pi i\nu(\beta+b)/j}, \text{ if } c\equiv 0,\ a\equiv d\equiv 1 \pmod{j}\\ &= \tfrac{1}{16}e^{-2\pi i\nu(a\beta+b)/j} = \tfrac{1}{16}e^{-2\pi i\nu(-b+\beta)/j}, \text{ if } c\equiv 0,\ a\equiv d\equiv -1 \pmod{j}\\ &= 0, \text{ otherwise.}\end{aligned}$$

Hence

$$(3.7)\qquad \delta_\nu(M)\,e^{-2\pi i\nu\beta/j} = \delta_\nu(MV).$$

Also, by (2.2), we have

$$(3.8)\qquad \begin{aligned}&e^{2\pi i m\beta/j}a_m(\nu,j,M)\\ &\qquad = \frac{\pi}{8j}\sum_{k=1}^{\infty}{}^{M}\,k^{-1}e^{2\pi i m\beta/j}A_{k,\nu}(m,M)\,(\nu/m)^{\frac{1}{2}}I_1\left(\frac{4\pi\sqrt{m\nu}}{jk}\right).\end{aligned}$$

But,

$$\begin{aligned}e^{2\pi i m\beta/j}A_{k,\nu}(m,M) &= \sum_{\substack{0\leqq h<jk\\(h,k)=1}}^{M}\left[-\frac{2\pi i}{jk}\{\nu h'+m(h-\beta k)\}\right]\\ &= \sum_{\substack{0\leqq h<jk\\(h,k)=1}}{}^{MV}\exp\left[-\frac{2\pi i}{jk}(\nu h'+mh)\right],\end{aligned}$$

as a direct consequence of (3.5), the summation restrictions (2.3), and the fact that when h is replaced by $h-\beta k$, k' is replaced by $k'+\beta h'\equiv a\beta+b \pmod{j}$. It now follows from (3.8) that

$$(3.9)\qquad e^{2\pi i m\beta/j}a_m(\nu,j,M) = a_m(\nu,j,MV).$$

Combining (3.6), (3.7), and (3.8), we have

$$\lambda_\nu(j,M;V\tau) = \lambda_\nu(j,MV;\tau),$$

and the proof is complete.

2. We will now apply Theorem (3.1) to show that the functions $\lambda_\nu(j,M;\tau)$ are abelian integrals of the second kind for $\Gamma(j)$. Let $P = r/s$,

$P'=r'/s'$ be two rational points on the real line with $(r,s)=(r',s')=1$ ($r/s=1/0=\infty$ is not excluded). Then P and P' are equivalent under $\Gamma(j)$ if and only if $r\equiv r'$, $s\equiv s' \pmod{j}$. Also in any fundamental domain $\mathfrak{F}_j$ the parabolic cusps are those (necessarily rational) points in which $\mathfrak{F}_j$ touches the real axis, including possibly the point at ∞. The *local uniformizing variable* at such a parabloic cusp is defined by (cf. [3, pp. 667-668])

$$t_P=\begin{cases}\exp\left[\dfrac{-2\pi i}{s^2 j(\tau-r/s)}\right], \text{ if } P=r/s<\infty,\ (r,s)=1\\ \exp(2\pi i\tau/j), \text{ if } P=\infty.\end{cases}$$

THEOREM (3.10). *The function* $\lambda_\nu(j,M;\tau)$ *is an abelian integral of the second kind for* $\Gamma(j)$ *with precisely one pole of order* ν *in* $\mathfrak{F}_j$. *More specifically,* $\lambda_\nu(j,M;\tau)$ *is regular in* $\operatorname{Im}(\tau)>0$ *and*

(i) *for* $V=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\Gamma(j)$ *and* $\operatorname{Im}(\tau)>0$

$$\begin{aligned}\lambda_\nu(j,M;V\tau)-\lambda_\nu(j,M;\tau)&=\sigma_\nu(j,M;V), \text{ if } \gamma\neq 0\\ &=0, \text{ if } \gamma=0,\end{aligned}$$

where $\sigma_\nu(j,M;V)$ *is defined by* (3.4),

(ii) *in* $\mathfrak{F}_j$ *there is precisely one parabolic cusp of the form* $P=-d_0/c_0$ *with* $(d_0,c_0)=1$ *and* $c_0\equiv c$, $d_0\equiv d \pmod{j}$, *where* $M=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$; *about this cusp* $\lambda_\nu(j,M;\tau)$ *has an expansion of the form*

$$b_{-\nu}t_P^{-\nu}+\sum_{m=0}^{\infty} b_m t_P^{m}, \tag{3.11}$$

(iii) *at any other cusp* P $\lambda_\nu(j,M;\tau)$ *has an expansion of the form*

$$\sum_{m=0}^{\infty} b_m t_P^{m}. \tag{3.12}$$

Proof. To obtain (i) we note that if $V\in\Gamma(j)$, $\lambda_\nu(j,MV;\tau)=\lambda_\nu(j,M;\tau)$. The case $\gamma=0$ is immediate; if $\gamma\neq 0$ we see from (3.3) that $\rho_\nu(j,M;V)=0$ for $V\in\Gamma(j)$. Hence by (3.2)

$$\lambda_\nu(j,M;V\tau)-\lambda_\nu(j,M;\tau)=\omega_\nu=\rho_\nu+\sigma_\nu=\sigma_\nu.$$

To prove (ii) and (iii) we suppose $P=r/s$, $(r,s)=1$, is a parabolic cusp of $\mathfrak{F}_j$, let $W=\begin{pmatrix}* & *\\ -s & r\end{pmatrix}\in\Gamma(1)$, and apply (3.2) with V replaced by $W^{-1}=\begin{pmatrix}r & *\\ s & *\end{pmatrix}$. This yields

$$\lambda_\nu(j,M;W^{-1}\tau)=\lambda_\nu(j,MW^{-1};\tau)+\omega_\nu(j,M;W^{-1}). \tag{3.13}$$

Now W maps $P = r/s$ onto ∞ and therefore, by (3.13) $\lambda_\nu(j, M; \tau)$ has the same behavior at r/s as $\lambda_\nu(j, MW^{-1}; \tau)$ has at ∞. More explicitly, the expansion of $\lambda_\nu(j, M; \tau)$ in powers of t_P contains a term of the form $t_P^{-\nu}$ when the expansion of $\lambda_\nu(j, MW^{-1}; \tau)$ in powers of $e^{2\pi i\tau/j}$ contains a term of the form $e^{-2\pi i\nu\tau/j}$, and otherwise contains only terms involving nonnegative powers of t_P.

However, the expansion of $\lambda_\nu(j, MW^{-1}; \tau)$ in powers of $e^{2\pi i\tau/j}$ is just the definition (2.4) of the function; we observe that this has a term of the form $e^{-2\pi i\nu\tau/j}$ if and only if $\delta_\nu(MW^{-1}) \neq 0$. Now

$$MW^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} r & * \\ s & * \end{pmatrix} = \begin{pmatrix} ar + bs & * \\ cr + ds & d_1 \end{pmatrix}$$

and by definiton of δ_ν, $\delta_\nu(MW^{-1}) \neq 0$ if and only if $ar + bs \equiv d_1 \equiv \pm 1$, $cr + ds \equiv 0 \pmod{j}$. Since the determinant of MW^{-1} is one, the condition $d_1 \equiv 1$ is a consequence of $cr + ds \equiv 0$, $ar + bs \equiv 1$, while $d_1 \equiv -1$ follows from $cr + ds \equiv 0$, $ar + bs \equiv -1$. Hence we see that $\delta_\nu(MW^{-1}) \neq 0$ if and only if $ar + bs \equiv \pm 1$, $cr + ds \equiv 0 \pmod{j}$. Since $ad - bc = 1$ these congruences hold if and only if $r \equiv d$, $s \equiv -c \pmod{j}$ or $r \equiv -d$, $s \equiv c \pmod{j}$. Thus $\delta_\nu(MW^{-1}) \neq 0$ if and only if P is of the form described in (ii) and the proof is complete.

It is worthwhile to make note of the fact that the proof of parts (ii) and (iii) above is much simpler than the proof of the corresponding result of [3]. This simplification has been made possible, oddly enough, by the greater generality of the functions we are treating here.

IV. Abelian integrals of the first kind. 1. We recall the assumption that ν is a positive integer and define $a_m(-\nu, j, M)$ to be the result of replacing ν by $-\nu$ everywhere in (2.2). That is,

$$(4.1) \qquad a_m(-\nu, j, M) = \left(\frac{\pi}{8j}\right) \sum_{k=1}^{\infty}{}^{M} k^{-1} A_{k,-\nu}(m, M) (-\nu/m)^{\frac{1}{2}} I_1\left(\frac{4\pi\sqrt{-m\nu}}{jk}\right),$$

with

$$A_{k,-\nu}(m, M) = \sum_{\substack{0 \leqq h < jk \\ (h,k)=1}}^{M} \exp\left[\frac{-2\pi i}{jk}(-\nu h' + mh)\right], \quad hh' \equiv -1 \pmod{k}.$$

We define $\delta_{-\nu}(M)$ in the same way from (2.1) and introduce the new functions

$$(4.2) \qquad \hat{\lambda}_{-\nu}(j, M; \tau) = \delta_{-\nu}(M) e^{2\pi i\nu\tau/j} + \sum_{m=1}^{\infty} a_m(-\nu, j, M) e^{2\pi i m\tau/j}.$$

In the terminology of [5] $\hat{\lambda}_{-\nu}$ is the *series supplementary* to λ_ν.

The importance of the supplementary series $\hat{\lambda}_{-\nu}$ resides in the fact that the proof of Theorem (3.1) in no way depends upon the positivity of ν. This can be seen by a careful examination of the proofs of the corresponding theorems in [2, 3]. Hence we obtain a result like Theorem (3.1) which, however, involves the function $\hat{\lambda}_{-\nu}$.

THEOREM (4.3). *Let* $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma(1)$. *Then, for* $\mathrm{Im}(\tau) > 0$, $\hat{\lambda}_{-\nu}(j, M; \tau)$ *is regular and satisfies*

$$\hat{\lambda}_{-\nu}(j, M; V\tau) = \hat{\lambda}_{-\nu}(j, MV; \tau) + \omega_{-\nu}(j, M; V), \tag{4.4}$$

where $\omega_{-\nu}(j, M; V)$ *is obtained from* $\omega_{\nu}(j, M; V)$ *by a replacement of* $-\nu$ *for* ν. *Hence we conclude that if* $\gamma = 0$, *then* $\omega_{-\nu}(j, M; V) = 0$, *and if* $\gamma \neq 0$, *then*

$$\omega_{-\nu}(j, M; V) = \rho_{-\nu}(j, M; V) + \sigma_{-\nu}(j, M; V),$$

where

$$\begin{aligned} \rho_{-\nu}(j, M; V) &= -\bar{\rho}_{\nu}(j, M; V) \\ \sigma_{-\nu}(j, M; V) &= \bar{\sigma}_{\nu}(j, M; V). \end{aligned} \tag{4.5}$$

Applying (4.4) and the reasoning used in the proof of Theorem (3.10), together with the fact that each $\hat{\lambda}_{-\nu}(j, M; \tau)$ is regular at ∞ (as is seen from (4.2)), we obtain

THEOREM (4.6). *The function* $\hat{\lambda}_{-\nu}(j, M; \tau)$ *is an abelian integral of the first kind for* $\Gamma(j)$. *Specifically, for* $\mathrm{Im}(\tau) > 0$ *and* $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma(j)$, *we have*

$$\begin{aligned} \hat{\lambda}_{-\nu}(j, M; V\tau) - \hat{\lambda}_{-\nu}(j, M; \tau) &= \sigma_{-\nu}(j, M; V), \text{ if } \gamma \neq 0 \\ &= 0, \text{ if } \gamma = 0, \end{aligned}$$

and at each parabolic cusp of $\mathcal{F}_j$, $\hat{\lambda}_{-\nu}(j, M; \tau)$ *has an expansion of the form* (3.12).

It is easy to see that certain linear combinations of the functions λ_{ν} are actually *modular functions* for $\Gamma(j)$. Such a linear combination can be constructed, for example, from any set of at least $q + 1$ of the λ_{ν}, where q is the minimal number of generators of $\Gamma(j)$. In this connection we have

THEOREM (4.7). *If* $\Lambda(\tau) = \sum_{M} \sum_{\nu} b(\nu, M) \lambda_{\nu}(j, M; \tau)$ *we let*

$$\hat{\Lambda}(\tau) = \sum_{M} \sum_{\nu} \bar{b}(\nu, M) \hat{\lambda}_{-\nu}(j, M; \tau).$$

Then $\Lambda(\tau)$ *is a modular function for* $\Gamma(j)$ *if and only if* $\hat{\Lambda}(\tau) \equiv 0$.

Proof. By Theorem (3.10), for $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma(j)$ and $\operatorname{Im}(\tau) > 0$.

$$\begin{aligned} \Lambda(V\tau) - \Lambda(\tau) &= \sum_{M} \sum_{\nu} b(\nu, M) \sigma_{\nu}(j, M; V), \text{ if } \gamma \neq 0 \\ &= 0, \text{ if } \gamma = 0. \end{aligned}$$

On the other hand Theorem (4.6) implies that

$$\begin{aligned} \hat{\Lambda}(V\tau) - \hat{\Lambda}(\tau) &= \sum_{M} \sum_{\nu} \bar{b}(\nu, M) \bar{\sigma}_{\nu}(j, M; V), \text{ if } \gamma \neq 0 \\ &= 0, \text{ if } \gamma = 0. \end{aligned}$$

Thus $\Lambda(\tau)$ is a modular function for $\Gamma(j)$ if and only if $\hat{\Lambda}(\tau)$ is. But we know that $\Lambda(\tau)$ is bounded in $\operatorname{Im}(\tau) > 0$, so that by a well known theorem, $\hat{\Lambda}(\tau)$ is a modular function if and only if it is a constant. By (4.2) that constant must be zero.

Remark. As a special case the theorem says that $\lambda_{\nu}(j, M; \tau)$ is a modular function if and only if $\hat{\lambda}_{-\nu}(j, M; \tau) \equiv 0$.

2. Our next objective is to show that finitely many of the $\hat{\lambda}_{-\nu}$ together with the constant function 1 can be chosen as a basis for the space of abelian integrals of the first kind. We will make crucial use of the following theorem from the classical theory of compact Riemann surfaces (cf. [15, p. 272]).

WEIERSTRASS GAP THEOREM. *Let S be a compact Riemann surface of genus g and let $P \in S$. Then there are g integers $0 < w_1 < w_2 < \cdots < w_g < 2g$, depending on P, such that there does not exist a meromorphic function on S having as its only singularity a pole of order w_i at P.*

For S we will take $\mathfrak{F}_j$ with appropriate sides identified; this is known to be a compact Riemann surface of genus

$$\begin{aligned} g &= 0, \text{ if } j = 2 \\ &= 1 + \frac{j^2(j-6)}{24} \prod_{p \mid j} \left(1 - \frac{1}{p^2}\right), \text{ if } j \geqq 3. \end{aligned}$$

The product is over all primes dividing j. It follows that the Riemann surface obtained from $\mathfrak{F}_j$ has genus zero exactly when $2 \leqq j \leqq 5$. We will refer to this Riemann surface simply as $\mathfrak{F}_j$. The meromorphic functions on $\mathfrak{F}_j$ are precisely the modular functions for $\Gamma(j)$.

We fix our attention on a single arbitrarily chosen parabolic cusp of $\mathfrak{F}_j$, say $P = r/s$, $(r, s) = 1$. Then there are integers t and u such that

$$M_P = \begin{pmatrix} t & u \\ s & -r \end{pmatrix} \in \Gamma(1);$$

by Theorem (3.10) $\lambda_\nu(j, M_P; \tau)$ has a pole of order ν at P and has no other singularity in $\mathfrak{F}_j$. We now consider the functions $\lambda_{w_i}(j, M_P; \tau)$, where $w_i = w_i(P)$ $(1 \leqq i \leqq g)$ forms the Weierstrass gap sequence at P, that is the sequence mentioned in the Weierstrass gap theorem.

Lemma (4.8). *No linear combination of the functions* $\lambda_{w_i}(j, M_P; \tau)$ *is a modular function for* $\Gamma(j)$ *except the trivial one with all coefficients zero.*

Proof. This is a direct application of the Weierstrass gap theorem since any such linear combination would be a meromorphic function on $\mathfrak{F}_j$ of the type excluded by that theorem.

Theorem (4.9). *The functions* $\hat{\lambda}_{-w_i}(j, M_P; \tau)$ $(1 \leqq i \leqq g)$ *are linearly independent and together with the constant function* 1, *they form a basis for the vector space of abelian integrals of the first kind for* $\Gamma(j)$.

Proof. If $\sum_{i=1}^{g} b_i \hat{\lambda}_{-w_i}$ is a constant, this constant must be zero by (4.2), and by Theorem (4.7), $\sum_{i=1}^{g} b_i \lambda_{w_i}$ is a modular function for $\Gamma(j)$. Hence by Lemma (4.8), $b_i = 0$ for $1 \leqq i \leqq g$, and we conclude that $1, \hat{\lambda}_{-w_1}, \cdots, \hat{\lambda}_{-w_g}$ are linearly independent. Now from the well known fact that vector space of abelian *differentials* on a compact Riemann has dimension equal to the genus (cf. [15, p. 252]), the remainder of the theorem follows.

3. Suppose $f(\tau)$ is any abelian integral of the second kind for $\Gamma(j)$. Then a linear combination of the functions $\lambda_\nu(j, M; \tau)$ can be so constructed that its principal part agrees with the principal part of $f(\tau)$ at each parabolic cusp. That is, there exists constants $b(\nu, M)$ such that

$$f(\tau) - \sum_M \sum_\nu b(\nu, M) \lambda_\nu(j, M; \tau) \tag{4.10}$$

is an abelian integral of the first kind. Thus we see that the functions λ_ν form a linear basis for the (infinite dimensional) vector space of abelian integrals of the second kind modulo the vector space of abelian integrals of the first kind.

Since (4.10) is an abelian integral of the first kind, by Theorem (4.9) there are constants b_i $(1 \leqq i \leqq g+1)$, depending on P of course, such that

$$f(\tau) = \sum_M \sum_\nu b(\nu, M) \lambda_\nu(j, M; \tau) + \sum_{i=1}^{g} b_i \hat{\lambda}_{-w_i}(j, M_P; \tau) + b_{g+1}. \tag{4.11}$$

We restate this in the following theorem.

Theorem (4.12). *The functions* $\lambda_\nu(j, M; \tau)$ $(M \in \Gamma(1)/\Gamma(j), \nu \geqq 1)$

together with the functions $1, \hat{\lambda}_{-w_1}(j, M_P; \tau), \cdots, \hat{\lambda}_{-w_g}(j, M_P; \tau)$ *form a basis*[8] *for the vector space of abelian integrals of the second kind for* $\Gamma(j)$. *Here* P *is any parabolic cusp of* $\mathfrak{F}_j$ *and* $w_1, \cdots, w_g$ *is the Weierstrass gap sequence at* P.

V. Modular functions. 1. Since a modular function $f(\tau)$ for $\Gamma(j)$ is an abelian integral of the second kind for $\Gamma(j)$, it follows that $f(\tau)$ can be expressed in the form (4.11). Moreover, in this case we can conclude that the constants $b_1, \cdots, b_g$ which occur in (4.11) are all zero. That is, we have

THEOREM (5.1). *If* $f(\tau)$ *is a modular function connected with* $\Gamma(j)$, *then*

$$f(\tau) = \sum_M \sum_\nu b(\nu, M) \lambda_\nu(j, M; \tau) + \textit{constant}, \tag{5.2}$$

where the linear combination on the right hand side is constructed so that its principal part agrees with that of $f(\tau)$ *at each parabolic cusp of* $\mathfrak{F}_j$.

Proof. By Theorems (3.10) and (4.6) we have for $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma(j)$ and $\operatorname{Im}(\tau) > 0$,

$$\begin{aligned} f(V\tau) - f(\tau) &= \sum_M \sum_\nu b(\nu, M) \sigma_\nu(j, M; V) + \sum_{i=1}^{g} b_i \bar{\sigma}_{w_i}(j, M_P; V), \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \gamma \neq 0 \\ &\quad 0, \text{ if } \gamma = 0. \end{aligned} \tag{5.3}$$

On the other hand, we consider the function

$$\hat{f}(\tau) = \sum_{i=1}^{g} \bar{b}_i \lambda_{w_i}(j, M_P; \tau) + \sum_M \sum_\nu \bar{b}(\nu, M) \hat{\lambda}_{-\nu}(j, M; \tau)$$

and again apply Theorems (3.10) and (4.6) to conclude that

$$\begin{aligned} \hat{f}(V\tau) - \hat{f}(\tau) &= \sum_{i=1}^{g} \bar{b}_i \sigma_{w_i}(j, M_P; V) + \sum_M \sum_\nu \bar{b}(\nu, M) \bar{\sigma}_\nu(j, M; V) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \gamma \neq 0 \\ &= 0, \text{ if } \gamma = 0. \end{aligned} \tag{5.4}$$

Now $f(\tau)$ is a modular function for $\Gamma(j)$ if and only if the right hand side of (5.3) is zero for all $V \in \Gamma(j)$. But this happens if and only if the right hand side of (5.4) is zero and therefore $f(\tau)$ is a modular function if and only if $\hat{f}(\tau)$ is a modular function.

---

[8] This is not a basis in the ordinary sense since several different $M \in \Gamma(1)/\Gamma(j)$ give rise to λ_ν's with poles at the same cusp, and therefore this set of functions is not linearly independent. However, restricting the choice of M in such a way that each curp occurs exactly once, we obtain a basis in the usual sense.

If there is an i with $1 \leqq i \leqq g$ and such that $b_i \neq 0$, then $\hat{f}(\tau)$ is a function on $\mathfrak{F}_j$ of the type excluded by the Weierstrass gap theorem. Hence if $f(\tau)$ is a modular function then $b_i = 0$ for $1 \leqq i \leqq g$, and we are done.

Remark. We can assert more than the statement of the theorem. For if $b_i = 0$, $1 \leqq i \leqq g$, then $\hat{f}(\tau)$ is a modular function which is bounded in the upper half-plane and must be a constant. From the definition of $\hat{f}(\tau)$, this constant must be zero. Conversely, if $\hat{f}(\tau) \equiv 0$, then $b_i = 0$ for $1 \leqq i \leqq g$ and $\sum_M \sum_\nu \bar{b}(\nu, M) \hat{\lambda}_{-\nu}(j, M; \tau) \equiv 0$. By Theorem (4.7) we conclude that $f(\tau) = \sum_M \sum_\nu b(\nu, M) \lambda_\nu(j, M; \tau)$ is a modular function. Thus we have shown that $f(\tau)$ *is a modular function if and only if* $\hat{f}(\tau) \equiv 0$, a generalization of Theorem (4.7).

From Theorem (5.1) and the definition (2.5) of $\lambda_\nu(j, M; \tau)$ we can now read off the Fourier coefficients of $f(\tau)$ in its expansion about ∞ (i.e. in powers of $e^{2\pi i\tau/j}$); the result is in terms of the principal part of $f(\tau)$ at each parabolic cusp of $\mathfrak{F}_j$. Specifically, if

$$f(\tau) = \sum_\nu \Big(\sum_M b(\nu, M) \delta_\nu(M) \Big) e^{-2\pi i\nu\tau/j} + \sum_{m=0}^{\infty} a_m e^{2\pi i m\tau/j},$$

then, for $m \geqq 1$,

$$(5.5) \qquad a_m = \frac{\pi}{8j} \sum_M \sum_\nu b(\nu, M) \sum_{k=1}^{\infty}{}^{M} k^{-1} A_{k,\nu}(m, M) (\nu/m)^{\frac{1}{2}} I_1\left(\frac{4\pi\sqrt{m\nu}}{jk}\right).$$

The expansion of $f(\tau)$ about each of the other cusps of $\mathfrak{F}_j$ can be obtained by applying Theorem (3.1), part (i), to $f(\tau)$.

It should be possible to obtain (5.5) by means of some variation of Rademacher's refinement of the circle method, introduced in [9] to obtain the Fourier coefficients of $J(\tau)$, the modular function for $\Gamma(1)$. In applying this method the nontrivial estimate (2.5) would be crucial, as it has been in our development here.

2. Theorem (5.1) does not say anything about the existence of modular functions for $\Gamma(j)$. We have indicated in § IV. 1 how modular functions can be constructed by forming suitable linear combinations of the functions λ_ν. We now sharpen this result considerably.

To begin with, since the functions λ_ν and $\hat{\lambda}_{-\nu}$ are abelian integrals of the second kind, they have no logarithmic singularities at the parabolic cusps of $\mathfrak{F}_j$ and thus they are invariant under transformation by *parabolic* $V \in \Gamma(j)$ (i. e. $V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ such that $|\alpha + \delta| = 2$). On the other hand it is known (cf. [1, p. 739]) that a set of generators for $\Gamma(j)$ can be found in which all but $2g$ are parabolic, where g is the genus of $\mathfrak{F}_j$. We label these nonparabolic generators $V_1, \cdots, V_{2g}$. Thus if an abelian integral of the second kind for $\Gamma(j)$ is invariant with respect to $V_1, \cdots, V_{2g}$ it is a modular function for $\Gamma(j)$.

Let $1 \leqq \nu_1 < \nu_2 < \cdots < \nu_n$ be integers. The linear combination

$$f(\tau) = \sum_{k=1}^{n} a_k \lambda_{\nu_k}(j, M; \tau) + \sum_{i=1}^{g} b_i \hat{\lambda}_{-w_i}(j, M_P; \tau),$$

with arbitrary $a_1, \cdots, a_n$, $b_1, \cdots, b_g$, is an abelian integral of the second kind satisfying

$$f(V_t \tau) = f(\tau) + \sum_{k=1}^{n} a_k \sigma_{\nu_k}(j, M; V_t) + \sum_{i=1}^{g} b_i \bar{\sigma}_{w_i}(j, M_P; V_t),$$

for $1 \leqq t \leqq 2g$. The homogeneous linear system

$$(5.6) \qquad \sum_{k=1}^{n} a_k \sigma_{\nu_k}(j, M; V_t) + \sum_{i=1}^{g} b_i \bar{\sigma}_{w_i}(j, M_P; V_t) = 0, \ 1 \leqq t \leqq 2g,$$

in the $n + g$ unknowns $a_1, \cdots, a_n$, $b_1, \cdots, b_g$ has $n - g$ linearly independent solutions, and in particular if $n \geqq g + 1$ it has a nontrivial solution. With $a_1, \cdots, a_n$, $b_1, \cdots, b_g$ a solution of (5.6) $f(\tau)$ is a modular function and therefore by the proof of Theorem (5.1), $b_1 = b_2 = \cdots = b_g = 0$. Thus some $a_k \neq 0$, and we have constructed a modular function for $\Gamma(j)$ whose only singularity is a pole of order at least ν_1 and at most ν_n at one of the parabolic cusps of $\mathfrak{F}_j$.

If we choose $n = g + 1$ and $\nu_k = k$, we have a nonconstant modular function for $\Gamma(j)$ whose only singularity is a pole of order at most $g + 1$. Note that the pole can be placed at any parabolic cusp of $\mathfrak{F}_j$. This is the best that can be done without some knowledge of the Weierstrass points of $\mathfrak{F}_j$[4] (cf. [15, pp. 272-274]).

VI. Concluding remarks. If Γ is any congruence subgroup of level j (i. e. if $\Gamma(j) \leqq \Gamma \leqq \Gamma(1)$), then any abelian integral of the second kind for Γ is, in particular, an abelian integral of the second kind for $\Gamma(j)$ and therefore can be expressed as in (4.11). This is not, however, a satisfactory situation since the functions λ_ν and $\hat{\lambda}_{-\nu}$ are not necessarily abelian integrals for Γ. A more satisfactory approach is to replace the functions λ_ν and $\hat{\lambda}_{-\nu}$ by similar functions whose construction is based upon the structure of Γ.

---

[4] It has been brought to my attention that B. Schoeneberg (Hamb. Abh. 17 (1951), 104-111) has shown that the parabolic cusps of $\mathfrak{F}_j$ are always Weierstrass points when ever $g \geqq 2$, that is, when $j \geqq 7$.

THE UNIVERSITY OF WISCONSIN,
MADISON, WISCONSIN.

REFERENCES.

[1] E. Hecke, *Mathematische Werke*, Göttingen, Vandenhoeck & Ruprecht, 1959.
[2] M. I. Knopp, "Construction of a class of modular functions and forms," *Pacific Journal of Mathematics*, vol. 11 (1961), pp. 275-293.
[3] ———, "Construction of a class of modular functions and forms, II," *Pacific Journal of Mathematics*, vol. 11 (1961), pp. 661-678.
[4] ———, "Fourier series of automorphic forms of nonnegative dimension," *Illinois Journal of Mathematics*, vol. 5 (1961), pp. 18-42.
[5] M. I. Knopp and J. Lehner, "On complementary automorphic forms and supplementary Fourier series," *Illinois Journal of Mathematics*, vol. 6 (1962), pp. 98-106.
[6] J. Lehner, "The Fourier coefficients of automorphic forms belonging to a class of horocyclic groups," *Michigan Mathematical Journal*, vol. 4 (1957), pp. 265-279.
[7] ———, "The Fourier coefficients of automorphic forms on horocyclic groups, II," *Michigan Mathematical Journal*, vol. 6 (1959), pp. 173-193.
[8] ———, "On modular forms of negative dimension," *Michigan Mathematical Journal*, vol. 6 (1959), pp. 71-88.
[9] H. Rademacher, "The Fourier coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 60 (1938), pp. 501-512.
[10] ———, "The Fourier series and the functional equation of the absolute modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 61 (1939), pp. 237-248.
[11] H. Rademacher and H. S. Zuckerman, "On the Fourier coefficients of certain modular forms of positive dimension," *Annals of Mathematics*, vol. 39 (1938), pp. 433-462.
[12] J. Raleigh, "The Fourier coefficients of the invariants $j(2^{\frac{1}{2}};\tau)$ and $j(3^{\frac{1}{2}};\tau)$," *Transactions of the American Mathematical Society*, vol. 87 (1958), pp. 90-107.
[13] W. H. Simons, "The Fourier Coefficients of the modular function $\lambda(\tau)$," *Canadian Journal of Mathematics*, vol. 4 (1952), pp. 67-80.
[14] J. R. Smart, "Modular forms of dimension -2 for subgroups of the modular group," Ph. D. thesis, Michigan State University, 1961.
[15] G. Springer, *Introduction to Riemann surfaces*, Reading, Addison-Wesley, 1957.
[16] H. S. Zuckerman, "On the coefficients of certain modular forms belonging to subgroups of the modular group," *Transactions of the American Mathematical Society*, vol. 45 (1939), pp. 298-321.

ON THE RITT POLYGON PROCESS.*

By Abraham P. Hillman and David G. Mead.

---

1. **Introduction.** The object of this paper is to make more accessible the technique of Ritt's polygon process together with those results in the theory of the components of the manifold of a partial differential polynomial that have been obtained with this technique.[1]

One of the authors has shown that the basic results of Ritt in the above mentioned theory follow from the Leading Coefficient Theorem, a generalization of Ritt's theorem on the terms of least degree.[2] No formal proof of the Leading Coefficient Theorem has previously been presented, although [1] indicates how the result could be obtained using Ritt's infinite series solutions. A proof which uses the essence of the polygon process but does not employ series solutions is the main part of this paper. The only previous material of differential algebra needed consists of conventions of notation[3] and of the two following results:

I. (Raudenbush). If A is a partial differential polynomial (p. d. p.) of positive degree in the ring $\mathcal{F}\{y_1, \cdots, y_n\}$ of p. d. p. in the y_i over a partial differential field $\mathcal{F}$, then there exists an extension $\mathcal{E}$ of $\mathcal{F}$ such that $\mathcal{E}$ contains a zero of A.

II. (Ritt's Lemma). Let A in $\mathcal{F}\{y_1, \cdots, y_n\}$ be of degree d in the $\delta_j y_i$ and let A be held by every partial derivative $\delta_j A$ of A with δ_j of order less than d.[4] Then $A = \alpha L^d$ with α in $\mathcal{F}$ and L linear in the $\delta_j y_i$.

Raudenbush gave a simple self-contained proof of (I) for ordinary differential polynomials in [2]; the proof applies readily in the partial case. In Appendix II below we give a formal proof of Ritt's Lemma essentially as it is established, without being formally stated, on p. 111 of [4].

---

* Received May 28, 1962.

[1] These results are mentioned in the introduction of [1]. (Numbers in square brackets refer to the bibliography.)

[2] See §§ 7-11 of [1] and Appendix I below.

[3] The differential notation of § 1 of [1] is used in this paper. The general Ritt theory of differential algebra is in [5].

[4] Ritt used "H holds A" to mean "H vanishes for the zeros of A" or in more technical language "H is in the perfect ideal generated by A."

2. The Leading Coefficient Theorem.

THEOREM. *Let*

$$B = M_0A + M_1(\delta_1A) + \cdots + M_s(\delta_sA) \tag{1}$$

where

$$A = A_0 + A_1c + \cdots + A_hc^h + \cdots; \ M_j = M_{j0} + M_{j1}c + \cdots$$

for $j = 0, 1, \cdots, s$; the A_h and M_{jh} are p. d. p. in $\mathfrak{R} = \mathfrak{F}\{y_1, \cdots, y_n\}$; and c is a transcendental constant over the partial differential field $\mathfrak{F}$. If A_0 is a non-zero element of $\mathfrak{R}$ which is annulled by $y_i = 0$ $(i = 1, \cdots, n)$ then the leading coefficient of c in B (i. e., the coefficient of the lowest power of c with a non-zero coefficient) is also annulled by $y_i = 0$.[5]

We assume the theorem to be false and obtain a contradiction with an inductive use of well ordering of pairs of non-negative integers.

Let a be the degree of the terms of least degree, in the δ_jy_i, of A_0. Let b be the power of c multiplying the leading coefficient of B. With each ordered pair (A, B) of series in c related as in (1), we associate the ordered pair (a, b) of non-negative integers. The (a, b) are well ordered by defining $(a, b) < (a', b')$ to mean that $a < a'$ or $a = a'$ and $b < b'$. If the theorem is false, there is a least ordered pair (a, b) for which some associated (A, B) satisfies the hypothesis but not the conclusion of the theorem. We now let (A, B) designate such a pair. Then B is of the form

$$B = B_bc^b + B_{b+1}c^{b+1} + \cdots$$

and there is a term in B_b free of the δ_jy_i since B_b is not annulled by $y_i = 0$. Since

$$B_b = \sum_{j=0}^{s} \sum_{h=0}^{b} M_{jh}(\delta_jA_{b-h}),$$

at least one of the A_h with $h \leqq b$ has a term free of the δ_jy_i. Denote by q the smallest h for which A_h has such a term. From the preceding observation and the fact that A_0 vanishes for $y_i = 0$, one notes that

$$0 < q \leqq b. \tag{2}$$

3. The cycle. Let t_h be the degree of the terms of least degree of A_h. (If $A_h = 0$ let $t_h = \infty$.) In particular, $t_0 = a$. Let p be the positive rational number defined by

$$p = \max[(q - h)/t_h], \ 0 \leqq h < q.$$

---

[5] It is easily seen that Theorem II of [1] is equivalent to this theorem.

If T is an arbitrary term of degree t of an A_h with $0 \leqq h < q$, then $p \geqq (q-h)/t_h \geqq (q-h)/t$, so that

$$h + pt \geqq q. \tag{3}$$

The substitution

$$y_i = c^p z_i \qquad (i = 1, \cdots, n), \tag{4}$$

where the z_i are new indeterminates, takes Tc^h into $T'c^{h+pt}$, where T' is a term of degree t in the $\delta_j z_i$. Let A_z and B_z be the results of applying (4) to A and B. In view of (3), it is clear that $A_z = \bar{A}_q c^q + \bar{A}_{q+1} c^{q+1} + \cdots$ and that the terms of $\bar{A}_q$ consist of a zero degree term from A_q and terms of positive degree from those A_h with $h < q$ for which $(q-h)/t_h = p$. The method of choosing p shows that there are such A_h with $h < q$. The terms of $\bar{A}_q$ from an A_h have degree $t_h = (q-h)/p$ in the $\delta_j z_i$. Obviously these t_h decrease as h increases; hence terms in $\bar{A}_q$ from an A_h do not cancel terms from an A_k with $h \neq k$. Let d be the degree in the $\delta_j z_i$ of $\bar{A}_q$. Then d is positive and

$$a \geqq d \tag{5}$$

since $a = t_0$, $d = t_h$ for some h with $0 \leqq h < q$, and the t_h decrease with increasing h.

Since the substitution (4) does not affect terms free of the indeterminates and raises the degree in c for all other terms, we have

$$B_z = \bar{B}_b c^b + \bar{B}_{b+1} c^{b+1} + \cdots,$$

where $\bar{B}_b$ is free of the $\delta_j z_i$.

Since $\bar{A}_q$ is of degree d, there are partial derivatives $\delta_j \bar{A}_q$ of $\bar{A}_q$, with δ_j of order d, which are non-zero elements of $\mathcal{F}$ and hence do not vanish for any zero of $\bar{A}_q$. The multiplicity of a zero $z_i = \theta_i$ $(i = 1, \cdots, n)$ of $\bar{A}_q$ is the minimum of the orders of the δ_j for which $z_i = \theta_i$ does not annul $\delta_j \bar{A}_q$. Let $z_i = \phi_i$ be a zero of $\bar{A}_q$ of lowest possible multiplicity and denote its multiplicity by f. Clearly

$$f \leqq d. \tag{6}$$

Now let $z_i = w_i + \phi_i$ where the w_i are new indeterminates. A_z and B_z go over into A_w and B_w which, as power series in c, start with terms of degree q and b respectively. (The differential field $\mathcal{F}$ may have to be extended to include the ϕ_i.) Let $F = A_w/c^q$ and $G = B_w/c^q$. Let $p = r/s$ with r and s

relatively prime positive integers. Substitute k for $c^{1/s}$, thus eliminating fractional powers of the constant, if there are any. Now

$$F = F_0 + F_1 k + F_2 k^2 + \cdots$$
$$G = G_g k^g + G_{g+1} k^{g+1} + \cdots$$

where $g = (b - q)s$, $G_g = \bar{B}_b$ is free of the $\delta_j w_i$, and the least degree terms of F_0 are of degree f (the multiplicity of $z_i = \phi_i$) in the $\delta_j w_i$.

4. Conclusion of the proof. The ordered pair (F, G) satisfies the hypothesis of the theorem (with c and the y_i replaced by k and the w_i). G_g does not vanish for $w_i = 0$ since it is a non-zero element of $\mathfrak{F}$. Using (5) and (6) above we have

$$f \leqq d \leqq a.$$

Hence $f \leqq a$. If $f < a$, the new ordered pair (f, g) precedes (a, b) and we have our contradiction. So let $f = a$. Now

$$f = d = a. \tag{7}$$

We consider two cases. First let p be an integer. Then the s in $p = r/s$ is 1, $g = (b - q)s = b - q$, and it follows from (2) that $g < b$. This means that $(f, g) = (a, g)$ precedes (a, b), which contradicts the minimal nature of (a, b).

We now need only consider the case in which p is not an integer. Then any t_h for which $p = (q - h)/t_h$ cannot be 1. Hence $\bar{A}_q$ does not have any terms of degree 1 in the $\delta_j z_i$. Since $\bar{A}_q$ has a zero degree term (from A_q) and has no first degree terms it is not of the form αL^d with α in $\mathfrak{F}$ and L linear in the $\delta_j z_i$. It follows from Ritt's Lemma that there is a zero of $\bar{A}_q$ of multiplicity less than d. This means that the multiplicity f of $z_i = \phi_i$ is less than d, contradicting (7). This completes the proof of the Leading Coefficient Theorem.

5. Relation to series solutions. The cycle described above indicates the method of obtaining the coefficients ϕ_i of a given power of c in a series solution for the A of (1).

Appendix I.

A major application of the Ritt polygon process is in the proof of Ritt's *Least Degree Theorem* which plays a fundamental role in the general theory. That the Least Degree Theorem follows easily from the Leading Coefficient

Theorem is essentially established in Example 6 of [1]. We now show this in a simple fashion by removing it from the context of [1].

THEOREM ON THE TERMS OF LEAST DEGREE. (Ritt). *Let* $F \neq 0$ *and* G *be p.d.p. in* $\mathfrak{R} = \mathfrak{F}\{y_1, \cdots, y_n\}$. *Let* F_1 *consist of the terms of* F *of lowest degree in the* $\delta_j y_i$ *and let* G_1 *be the similar part of* G. *If* G *holds* F, *then* G_1 *holds* F_1.

Proof. Given that G holds F, we have

$$G^t = \sum_{h=0}^{s} M_h(\delta_h F) \tag{1'}$$

with M_i in $\mathfrak{R}$ and t a positive integer. The terms of least degree of G^t are $(G_1)^t$. We make the substitution $y_i = cz_i$ where the z_i are new indeterminates and c is a transcendental constant over $\mathfrak{F}$. Then F and G^t go over into (finite) series in c with leading coefficients that are F_1 and $(G_1)^t$ respectively with the y_i replaced by the z_i.

The theorem is obvious if F_1 has no zeros. So let $y_i = \phi_i$ $(i = 1, \cdots, n)$ be a zero of F_1 and let us substitute $z_i = w_i + \phi_i$, where the w_i are new indeterminates. The two substitutions send F and G into $\bar{F}$ and $\bar{G}$ and replace equation (1′) by an equation similar to equation (1) in the Leading Coefficient Theorem above. The leading coefficient of $\bar{F}$ vanishes for $w_i = 0$ and hence so does that of $(\bar{G})^t$. Therefore $(G_1)^t$ and also G_1 vanishes for $y_i = \phi_i$, completing the proof.

Appendix II.

We present here a formal treatment of some of Ritt's material on p. 111 of [4].

RITT'S LEMMA. *Let* $A \in \mathfrak{R} = \mathfrak{F}\{y_1, \cdots, y_n\}$. *Let* A *be of degree* d *in the* $\delta_j y_i$ *and let* A *be held by every partial derivative* $\delta_j A$ *of* A *with* δ_j *of order less than* d. *Then* $A = \alpha L^d$ *where* $\alpha \in \mathfrak{F}$ *and* L *is a linear element of* $\mathfrak{R}$.

Proof. Let L be a non-zero partial derivative δA of A with δ of order $d - 1$. Then L is linear and involves at least one of the y_i, say y_1. Let W be a factor of A which is algebraically irreducible in $\mathfrak{R}$ and in which as high[6] a derivative of y_1 is present as the highest in A. Now L holds W since it holds A. From the nature of the $\delta_j y_i$ in W, this implies that $L^t = BW$ with t a positive integer and $B \in \mathfrak{R}$. The irreducibility of W and the linearity of L imply that $W = \gamma L$ with $\gamma \in \mathfrak{F}$. Thus any factor of A differing from L

---

[6] This refers to the customary type of ordering as, for example, in [3].

has only lower derivatives of y_1 than the highest in A (or L). Let $A = QL^e$ with Q not a multiple of L. Then there is an e-th order derivative of A of the form βQ with $\beta \in \mathfrak{F}$. If $e < d$, the hypothesis of this lemma tells us that Q holds A. This is not possible due to the nature of the $\delta_j y_i$ in Q. Hence $e = d$ and $Q \in \mathfrak{F}$ as desired.

UNIVERSITY OF SANTA CLARA AND PRATT INSTITUTE.

REFERENCES.

[1] A. P. Hillman, "On the differential algebra of a single differential polynomial," *Annals of Mathematics*, vol. 56 (1952), pp. 157-168.

[2] H. W. Raudenbush, "Differential fields and ideals of differential forms," *Annals of Mathematics*, vol. 34 (1933), pp. 509-517.

[3] ———, "Hypertranscendental adjunctions to partial differential fields," ***Bulletin of the American Mathematical Society***, vol. 40 (1934), pp. 714-720.

[4] J. F. Ritt, "On the manifolds of partial differential polynomials," *Annals of Mathematics*, vol. 46 (1945), pp. 102-112.

[5] ———, *Differential algebra*, American Mathematical Society Colloquium Publications, vol. 33.

THE AUTOMORPHISMS OF $PO_8^+(Q)$ AND $PS_8^+(Q)$.*

By MARÍA J. WONENBURGER.[1]

The purpose of the present paper is to complete some results of [8] by studying the automorphisms of $PS_8^+(Q)$, the projective group of proper similitudes of an 8-dimensional vector space M with respect to a non-degenerate quadratic form Q, and of its subgroup $PO_8^+(Q)$, the projective group of rotations. It is well-known that in certain cases these groups actually have exceptional automorphisms, that is, automorphisms which are not induced by automorphisms of $S_8^+(Q)$, the group of proper similitudes.

We prove, under the assumption that the base field K has characteristic not 2, that $PS_8^+(Q)$ has exceptional automorphisms if, and only if, either the quadratic form Q or a scalar multiple αQ permits composition. As for $PO_8^+(Q)$, it has exceptional automorphisms if, and only if, the vector space M has an orthonormal basis with respect to either Q or a multiple αQ, and the base field is Pythagorean; this implies $PO_8^+(Q) \cong PS_8^+(Q)$.

In the last section we show that, if $PS_8^+(Q)$ has exceptional automorphisms, any exceptional one together with the ones induced by the automorphisms of $S_8^+(Q)$ generate the whole group of automorphisms.

1. We are going to use the notation of [8, Sections 1, 5] with the only difference that now *Q is always a non-degenerate quadratic form over an 8-dimensional vector space M,* and we assume again that K has characteristic not 2. Moreover, since two quadratic forms which are obtained one from the other by scalar multiplication define the same group of similitudes and related groups we can assume in our arguments without loss of generality that there exists a vector $x \in M$ such that $Q(x) = 1$.

THEOREM 1. *The group $PS^+(Q)$ has exceptional automorphisms if, and only if, the quadratic form Q (up to a scalar factor) permits composition* (see e.g. [4]).

Proof. If an automorphism ϕ of $PS^+(Q)$ takes any $(6,2)$ coset into another $(6,2)$ coset, then the restriction of ϕ to $PO^+(Q)$ is an automorphism of $PO^+(Q)$ induced by an automorphism of $O^+(Q)$ (see [2, Section 37]) and

---

* Received July 9, 1962.

[1] Postdoctorate Fellow (of the National Research Council of Canada) at Queen's University.

therefore ϕ is an automorphism of $PS^{+}(Q)$ induced by an automorphism of $S^{+}(Q)$ (see the proof of [8, Lemma 2]).

We consider now two different cases.

Case 1. *K has more than* 5 *elements.* Then ϕ is an exceptional automorphism if, and only if, there exists a $(6,2)$ coset $\bar{U}$ which is not taken into a $(6,2)$ coset and therefore by [8, Corollary of Lemma 3] $\phi(\bar{U})$ is the coset of a P-involution. Let x_1, x_2 be any orthogonal basis for the minus-space of $\bar{U}$ and $x_3, x_4, \cdots, x_8$ any orthogonal basis of its plus-space. Let U_i, $i=2,3,\cdots,8$, be the $(6,2)$ involutions whose minus-spaces are spanned by x_1 and x_i, so that $\bar{U}_2=\bar{U}$. If $\phi(\bar{U}_i)=\bar{T}_i$, in the course of the proof of [8, Lemma 6] it was shown that the T_i, $i=2,3,\cdots,8$, anticommute with each other and, since the $\bar{T}_i$ are either $(6,2)$ cosets or cosets of P-involutions, the only possibility is that the T_i are anticommuting P-involutions.

Assume now that ρ and ρ' are the ratios of any two of these P-involutions, say T_2 and T_3, and let x be any vector of M such that $Q(x)=1$. The vectors x, xT_2, xT_3, xT_2T_3, are mutually orthogonal non-isotropic vectors (cf. [8, proof of Lemma 4]). Moreover, if N is the 4-dimensional space spanned by these vectors, N is invariant under the P-involutions T_2 and T_3; hence $N^{\perp}$, the orthogonal complement of N, is also invariant under T_2 and T_3. Now, if $y\in N^{\perp}$ and $Q(y)=\beta\neq 0$, y, yT_2, yT_3, yT_2T_3 is an orthogonal basis of $N^{\perp}$. We have found, then, an orthogonal basis, x, xT_2, xT_3, xT_2T_3, y, yT_2, yT_3, yT_2T_3, of M such that the lengths $Q(.)$ of its vectors are

$$1,\ \rho,\ \rho',\ \rho\rho',\ \beta,\ \beta\rho,\ \beta\rho',\ \beta\rho\rho',$$

respectively, which shows that the quadratic form Q permits composition (see [4, Sections 1, 2]). Hence, if $PS^{+}(Q)$ has exceptional automorphisms, the quadratic form Q (up to a scalar factor) is the quadratic form associated to a Cayley algebra.

On the other hand, it is mentioned in [1] that if Q is the quadratic form associated to a Cayley algebra, the automorphisms ϕ_1 and ϕ_2 of $PS^{+}(Q)$ defined in [1, Cor. 2 of Th. 1] are exceptional. (In the proof of Th. 3 below we show that the automorphism ϕ_2 is exceptional.)

Case 2. *K is a finite field.* If the quadratic form Q has square discriminant, Q is the quadratic form of a split Cayley algebra and the automorphisms ϕ_2 of [1, Cor. 2 of Th. 1] is exceptional. When the discriminant of Q is not a square Q does not permit composition and the commutator group of $PS^{+}(Q)$ is $PO^{+}(Q)\cong P\Omega(Q)$. This implies that any automorphism ϕ of $PS^{+}(Q)$ induces an automorphism in $PO^{+}(Q)$, but the automorphisms of

$PO^+(Q)$ are induced by automorphisms of $O^+(Q)$ (see [8, Th. 4]) and, consequently, ϕ is induced by an automorphism of $S^+(Q)$.

LEMMA 1. *If the group $PO^+(Q)$ has exceptional automorphisms the space M has an orthonormal basis with respect to Q or αQ* (see [2, p. 60]).

Proof. When K is a finite field, $PO^+(Q)$ does not have exceptional automorphisms ([8, Th. 4]). It follows from [7] that if K has more than 5 elements an automorphism of $PO^+(Q)$ can not take a $(6,2)$ coset into a $(4,4)$ coset. (The proof is like the proof of [8, Lemma 3] using $PO^+(Q)$ instead of $PS^+(Q)$.) Hence, under an automorphism ϕ of $PO^+(Q)$, a $(6,2)$ coset must go either into a $(6,2)$ coset or into the coset of an orthogonal P-involution. We have again that, if ϕ is an exceptional automorphism, there should exist a $(6,2)$ coset $\bar{U}$ which is taken by ϕ into the coset of an orthogonal P-involution, that is, a P-involution of ratio 1. Define $x_1, x_2, x_3, \cdots, x_8$ and the U_i as before and let $\phi(\bar{U}_i) = \bar{T}_i$. Then [8, Lemma 6] still applies and we can conclude that the T_i, $i = 2, 3, \cdots, 8$, are seven orthogonal P-involutions such that any two of them anticommute. If x is any vector with $Q(x) = 1$, then x together with the vectors xT_i, $i = 2, 3, \cdots, 8$ forms an orthonormal basis of M.

LEMMA 2. *If an automorphism of $PO^+(Q)$ (or $PS^+(Q)$, if K has more than 5 elements) is exceptional it takes every $(6,2)$ coset into the coset of a P-involution.*

Proof. If K is a finite field, $PO^+(Q)$ has no exceptional automorphism and there is nothing to prove. Hence we can assume that K has more than 5 elements. Suppose then that the automorphism ϕ takes the coset of a $(6,2)$ involution U into the coset of a P-involution. We have already shown that if x_1, x_2 and $x_3, \cdots, x_8$ are any orthogonal bases of the minus-space M_1 and plus-space $M_1{}^\perp$, respectively, of U, the cosets defined by the $(6,2)$ involutions U_i, $i = 2, \cdots, 8$, are taken into the cosets of anticommuting P-involutions T_i. Now, if U_{ij} is the involution whose minus-space is spanned by x_i and x_j, $i, j \neq 1$, $U_{ij} = U_i U_j$, hence $\phi(\bar{U}_{ij}) = \bar{T}_i \bar{T}_j$ and, as a consequence of $T_i T_j = -T_j T_i$, $T_i T_j$ is also a P-involution. This implies that any $(6,2)$ coset whose minus space is orthogonal to M_1 is mapped by ϕ into a P-involution.

Now, let N be the minus space of any $(6,2)$ involution V. The space $(M_1 + N)^\perp$ has dimension $\geqq 4$ and, since $(M_1 + N) \cap (M_1 + N)^\perp$ is at most 2 dimensional, $(M_1 + N)^\perp$ contains a 2-dimensional non-isotropic subspace R. Since M_1 is orthogonal to R, ϕ maps the coset of the $(6,2)$ involution with minus-space R into the coset of a P-involution, and therefore ϕ also maps $\bar{V}$ into the coset of a P-involution because R if orthogonal to N.

LEMMA 3. *If the vector space M has an orthonormal basis with respect to Q, then any orthogonal P-involution belongs to the commutator group of $O^+(Q)$.*

Proof. Since any P-involution takes any vector x into a vector orthogonal to x, it is an immediate consequence of Witt's theorem that given an orthogonal P-involution T it is possible to choose an orthonormal basis of M of the form $x_i, y_i = x_iT$, $i = 1, 2, 3, 4$ (cf. the proof of Lemma 4 below). Define the linear transformations T_1 and S as follows,

$$\begin{array}{lll} x_jT_1 = y_j, & y_jT_1 = -x_j, & \text{for } j = 1, 2 \\ x_jT_1 = x_j, & y_jT_1 = y_j, & \text{for } j = 3, 4, \text{ and} \\ x_iS = y_{i+2} & y_iS = x_{i+2}. & \text{where the indexes } i + 2 \text{ should} \end{array}$$

be computed modulo 4.

Then T_1 and S are rotations and $T = T_1ST_1^{-1}S_1^{-1}$.

THEOREM 2. *The group $PO^+(Q)$ has exceptional automorphisms if, and only if, M has an orthonormal basis with respect to Q or αQ and the base field K is Pythagorean.*

Proof. If $PO^+(Q)$ has exceptional automorphisms, let x_i, $i = 1, 2, \cdots, 8$ be an orthonormal basis of M with respect to Q or αQ. If K is not Pythagorean, assume that $\alpha^2 + \beta^2$ is not a square and let U be the $(6, 2)$ involution whose minus-space is spanned by x_1 and $\alpha x_2 + \beta x_3$. Then the coset of U does not belong to $P\Omega(Q)$ since the spin-norms of U and $-U$ belong to the quadratic class of $\alpha^2 + \beta^2$ which is not a square. Hence $\bar{U}$ can not be mapped by an automorphism of $PO^+(Q)$ into the coset of an orthogonal P-involution. Therefore, by Lemma 2, $PO^+(Q)$ has no exceptional automorphism. This proves the "only if" part of the theorem.

On the other hand, if the conditions of the theorem are satisfied, the ratio of any similitude is a square and, consequently, $PO^+(Q) \cong PS^+(Q)$. Hence the "if part" of the theorem follows from Theorem 1.

COROLLARY. *If $PO^+(Q)$ has exceptional automorphisms, then $PO^+(Q) \cong PS^+(Q)$.*

2. In this section we are going to determine the form of all the auto-

morphisms of $PS^+(Q)$. By the Corollary that we have just established, this gives us also the automorphisms of $PO^+(Q)$.

LEMMA 4. *In $PS^+(Q)$ the cosets of P-involutions of the same ratio decompose into two conjugate classes.*

Proof. We will prove first that in $S(Q)$ any two P-involutions T_1 and T_2 of ratio ρ are conjugate.

Let x_i, x_iT_1, $i=1,2,3,4$, be an orthogonal basis of M with respect to Q. Let y_1 be any vector such that $Q(y_1)=Q(x_1)$. Then the 2-dimensional subspaces N_1 and M_1 spanned by y_1 and y_1T_2, and x_1 and x_1T_1, respectively, are isometric. Hence, by Witt's theorem, $N_1{}^\perp$ and $M_1{}^\perp$ are also isometric and therefore there exists a vector $y_2\in N_1{}^\perp$ such that $Q(y_2)=Q(x_2)$. Now the subspaces N_2 and M_2 spanned by y_i, y_iT_2, $i=1,2$, and x_i, x_iT_1, $i=1,2$, respectively, are isometric and we can repeat the preceding process until we get an orthogonal basis y_i, y_iT_2, $i=1,2,3,4$ such that $Q(x_i)=Q(y_i)$. Define the linear transformation U by

$$x_iU=y_i, \qquad (x_iT_1)U=y_iT_2, \qquad i=1,2,3,4.$$

Then U is an orthogonal transformation and $T_2=U^{-1}T_1U$.

$PS(Q)$ consists of the subgroup $PS^+(Q)$ and the cosets $PS^-(Q)$, therefore in $PS^+(Q)$ the cosets defined by all the P-involutions of ratio ρ are either all conjugate or they decompose into the two conjugate classes. But, if $V\in S^-(Q)$, the coset defined by VTV^{-1} can not be equal to a coset defined by WTW^{-1} with $W\in S^+(Q)$, because this would imply

$$(W^{-1}V)T(W^{-1}V)^{-1}=\pm T, \qquad W^{-1}V\in S^-(Q),$$

and there are no improper similitudes commuting or anticommuting with T (see [6, Prop. 3 and 4]). This proves the lemma.

THEOREM 3. *If the quadratic form Q permits composition, then the group of automorphisms of $PS^+(Q)$ is generated by any exceptional automorphism and the ones induced by automorphisms of $S^+(Q)$.*

Proof. We are going to show first that the automorphism ϕ_2 of $PS^+(Q)$ defined in [1, Corollary 2 of Th. 1] is exceptional. Let e be the element of M which acts as the identity of the Cayley algebra and let b be any non-isotropic vector orthogonal to e. The proof of the corollary mentioned above shows that ϕ_2 takes the coset of the $(6,2)$ involution whose minus space is spanned by e and b into the coset of the transformation R_b defined by right multiplication by b. Hence R_b has ratio $Q(b)$ and $R_b{}^2$ is the multiplication by $-Q(b)$. This shows that R_b is a P-involution of ratio $Q(b)$ and ϕ_2 is an exceptional automorphism.

If there exists in $PS^+(Q)$ a P-involution T of ratio ρ, since eT is orthogonal to e and $Q(eT) = \rho$, the coset of the $(6,2)$ involution U whose minus-space is spanned by e and eT is taken by ϕ_2 into the coset of the P-involution R_{eT} of ratio ρ. Moreover the $(6,2)$ cosets conjugate to $\bar{U}$ are mapped by ϕ_2 into cosets conjugate to $\bar{R}_{eT}$ in $PS^+(Q)$.

We consider again two different cases.

Case 1. *K has more than 5 elements.* Let ψ be any exceptional automorphism of $PS^+(Q)$, $\bar{U}$ any $(6,2)$ coset and $\psi(\bar{U}) = \bar{T}$. Then, by Lemma 2, we know that T is a P-involution and we have two possibilities,

a) $\bar{T}$ is conjugate to $\bar{R}_{cT}$. Then $\phi_2{}^{-1}\psi$ takes $\bar{U}$ into another $(6,2)$ coset and therefore is an automorphism induced by an automorphism of $S^+(Q)$, that is, $\phi_2{}^{-1}\psi = \sigma_A$, where $\sigma_A(\bar{X}) = \overline{AXA^{-1}}$ and A is a semi-similitude, and $\psi = \phi_2\sigma_A$,

b) $\bar{T}$ is not conjugate to $\bar{R}_{eT}$. Then, if $V \in S^-(Q)$, $\overline{VTV^{-1}}$ is conjugate to $\bar{R}_{eT}$ and consequently σ_V, where σ_V is the automorphism of $PS^+(Q)$ defined by $\sigma_V(\bar{X}) = \overline{VXV^{-1}}$, is of the form $\sigma_V\psi = \phi_2\sigma_A$, that is, $\psi = \sigma_{V^{-1}}\phi_2\sigma_A$.

We have shown then that ϕ_2 and the automorphisms of the form σ_A generate the whole group of automorphisms and that the group generated by any exceptional automorphism ψ and the σ_A contains ϕ_2 and therefore is the whole group of automorphisms.

Case 2. *K is a finite field.* Then the second commutator group of $PS^+(Q)$ coincides with $P\Omega(Q)$, the commutator group of $PO^+(Q)$ (actually $P\Omega(Q)$ is the first commutator group of $PS^+(Q)$ by [5, Th. 5]). Hence any automorphism ϕ of $PS^+(Q)$ induces an automorphism in $P\Omega(Q)$ and it is known that in $P\Omega(Q)$ the $(6,2)$ cosets can not be taken into $(4,4)$ cosets (see [2, Section 43]) and therefore ϕ takes $(6,2)$ cosets either into $(6,2)$ cosets or into cosets of P-involutions. But the $(6,2)$ involutions whose cosets belong to $P\Omega(Q)$ are those whose minus-space has square discriminant with respect to the restriction of Q to it, and all such involutions are conjugate in $P\Omega(Q)$. Hence we have two possibilities, under ϕ

a) either all the $(6,2)$ cosets of $P\Omega(Q)$ go into $(6,2)$ cosets, in which case the restriction of ϕ to $P\Omega(Q)$ is an automorphism induced by an automorphism of $O^+(Q)$ (see [2, Sections 43, 42]) and then, as in the proof of [8, Lemma 2], we can conclude that ϕ itself is induced by an automorphism of $S^+(Q)$.

b) all the $(6,2)$ cosets of $P\Omega(Q)$ go into P-involutions. Then, as in case 1, we can deduce that either $\phi_2^{-1}\phi$ or $\phi_2^{-1}\sigma\gamma\phi$ is an automorphism induced by an automorphism of $S^+(Q)$, which proves the theorem.

Remark. To carry out some proofs we have considered two different cases. This is due to the fact that [8, Cor. of Lemma 3] was established under the condition that K has more than 5 elements, but, in fact, the corollary can be easily proved when K is the field with 5 elements. We have not proved it here because we would still need a different proof for the case of a field with 3 elements.

QUEEN'S UNIVERSITY, KINGSTON, ONTARIO.

REFERENCES.

[1] F. van der Blij and T. A. Springer, "Octaves and Triality," *Niew Archief voor Wiskunde* (3) 1960, pp. 158-169.

[2] J. Dieudonné, "On the automorphism of the classical groups," *Memoirs of the American Mathematical Society*, No. 2 (1951).

[3] ———, *La Géométrie des Groupes Classiques* (Berlin: Springer-Verlag, 1955).

[4] N. Jacobson, "Composition algebras and their automorphisms," *Rendiconti Circolo Matematico di Palermo* (II), vol. 7 (1958), pp. 55-80.

[5] M. J. Wonenburger, "The Clifford algebra and the group of similitudes," *Canadian Journal of Mathematics*, vol. 14 (1962), pp. 45-59.

[6] ———, "Study of certain similitudes," *Canadian Journal of Mathematics*, vol. 14 (1962), pp. 60-68.

[7] ———, "The automorphisms of the group of rotations and its projective group corresponding to quadratic forms of any index," *Canadian Journal of Mathematics*, 1963.

[8] ———, "The automorphisms of the group of similitudes and some related groups." *American Journal of Mathematics*, vol. 84 (1962), pp. 600-614.

FURTHER PATHOLOGIES IN ALGEBRAIC GEOMETRY.*[1]

By David Mumford.

The following note is not strictly a continuation of our previous note [1]. However, we wish to present two more examples of algebro-geometric phenomena which seem to us rather startling. The first relates to characteristic p behaviour, and the second relates to the hypothesis of the completeness of the characteristic linear system of a maximal algebraic family. We will use the same notations as in [1].

I.

The first example is an illustration of a general principle that might be said to be indicated by many of the pathologies of characteristic p:

A non-singular characteristic p variety is analogous to a general non-Kähler complex manifold; in particular, a projective embedding of such a variety is not as "strong" as a Kähler metric on a complex manifold; and the Hodge-Lefschetz-Dolbeault theorems on sheaf cohomology break down in every possible way.

In this case we wish to look at the two dimensional cohomology of an algebraic surface F, non-singular, and of any characteristic but 0. The surface we shall choose will (a) be specialization of a characteristic 0 surface F', and (b) will satisfy $q = h^{0,1} = h^{1,0}$. Consequently the second Betti number B_2 is the same, whether defined (i) as that of F' in the topological sense, (ii) as $h^{2,0} + h^{1,1} + h^{0,2}$, or (iii) following Igusa [2], as $\mathrm{Deg}(c_2) + 4q - 2$. Let ρ be the base number of F. Igusa showed that, in fact, $B_2 \geqq \rho$. However, in characteristic 0, one has the stronger result, $B_2 = h^{2,0} + h^{1,1} + h^{0,2} \geqq 2p_g + \rho$ (where $p_g = h^{2,0} = h^{0,2}$ is the geometric genus of F) as a result of the Hodge-Dolbeault theorems. Therefore the question arises whether this stronger inequality is valid in characteristic p. The answer is no.

A rather complicated example was discovered in 1961 by J. Tate and A. Ogg. Here is a very simple example: let E be a super-singular elliptic

---

* Received July 30, 1962.

[1] This work was supported by the Army Research Office (Durham).

curve of characteristic p (i.e. such that the rank of $End(E)$ is 4). Let $F = E \times E$. In this case, in fact:

$$\rho = B_2 = 6; \quad p_g = 1.$$

Here $p_g = 1$ since the sheaf $\Omega_F{}^2 \cong \mathcal{O}_F$; and $B_2 = 6$ by Igusa's definition, for example, since $\operatorname{Deg}(c_2) = 0$, and $q = 2$. Finally $\rho = 6$ since in general, for any two elliptic curves E_1 and E_2, one knows that the base number ρ for $E_1 \times E_2$ equals 2 plus the rank of $\operatorname{Hom}(E_1, E_2)$.

There remains one outstanding conjecture still neither proven nor disproven in characteristic p, which according to the general principle mentioned above ought to be false. This is the Regularity of the Adjoint, which may be stated as follows: if V is a non-singular projective surface and if H is a non-singular hyperplane section, then

$$H^1(\mathcal{O}_V) \rightarrow H^1(\mathcal{O}_H)$$

is injective.

II.

The second example concerns space curves in characteristic 0. Let A be any family of non-singular space curves, and let $a \in A$ represent the curve $\gamma \subset P_3$. Let T_a denote the Zariski tangent space to A at a, and let N denote the sheaf of sections of the normal bundle to γ in P_3. Then it is well-known [3] that there is a "characteristic" map:

$$T_a \rightarrow H^0(N).$$

The problem of completeness consists in asking when, for given γ, there is a family A containing γ such that the characteristic map is surjective. Kodaira [3] has shown that such a family exists if $H^1(N) = (0)$. Our example shows that if $H^1(N) \neq (0)$, then there need not be such a family.

In fact, in our example, this incompleteness holds for every curve in an open set of the corresponding Chow variety. Consequently, it is also an example where the Hilbert scheme [4] has a multiple component, i.e. is not reduced at one of its generic points. Another corollary of this example is obtained by blowing up such a space curve $\gamma \subset P_3$ to a surface E in a new three-dimensional variety V_3. Then Kodaira [5] has shown essentially that the local moduli scheme of the variety V_3 is isomorphic to the germ of the

Hilbert scheme of P_3 at the point corresponding to γ. Therefore we have constructed a non-singular projective three-dimensional variety whose local moduli scheme is nowhere reduced; in other words, any small deformation of V_3 is a variety the number of whose moduli is less than the dimension of $H^1(\Theta)$ (where Θ is the sheaf of vector fields).

The curves γ that we have in mind have degree 14 and genus 24. In the following, h will stand for the divisor class on γ induced by plane sections, and K_γ will stand for the canonical divisor on γ; also F will stand for a cubic or quartic surface in P_3, and H will stand for the (Cartier) divisor class on F induced by plane sections. The first step is partial classification of all space curves of this degree and genus, which confirms the results of M. Noether's well-known table [6].

(A) *Any non-singular space curve γ of degree 14 and genus 24 is contained in a pencil P of quartic surfaces.*

Proof. Since $\mathrm{Deg}(4h) = 56$, and $\mathrm{Deg}(K_\gamma) = 46$, the linear system $|4h|_\gamma$ is non-special,[2] and has dimension $56 - 24 = 32$. Since there is a 34-dimensional family of quartics in P_3, (A) follows.

There are 2 cases: (a) the pencil has no fixed components, and (b) the pencil has fixed components. In case (a), note first that if F' and F'' span P, then $F' \cdot F'' = \gamma + c$, where c is a conic. Now c has at most double points, hence $\gamma + c$ has at most triple points. Therefore no point x is a double point for both F' and F''. Noting that both F' and F'' are non-singular and transversal along $\gamma - c$, hence at all but a finite number of points of γ, it follows that almost every $F \in P$ is non-singular everywhere along γ.

(B) *Every algebraic family of space curves of type* (a) *has dimension less than or equal* 56.

Proof. It is enough to show that every family of pairs (γ, F) consisting of such curves γ, and quartics $F \supset \gamma$, F being non-singular along γ, has dimension at most 57. Now since all such quartics contains conics, they are not generic [7], and there is at most a $34 - 1 = 33$ dimensional family of quartics F involved in such a family of pairs. Moreover, the dimension of the set of all γ on one such F can be computed from the Riemann-Roch theorem on F:

---

[2] Here and below, $|D.|_V$ always means the linear system on V in which the Cartier divisor D varies. Also, $(D^2)_V$ always denotes the self-intersection of D, as a divisor class on D (assuming D effective).

$$\dim|\gamma|_F = \frac{\mathrm{Deg}(\gamma^2)_F}{2} + 1 + \{\dim H^1(\mathcal{O}_F(\gamma)) - \dim H^2(\mathcal{O}_F(\gamma))\}.$$

But $(\gamma^2)_F \equiv K_\gamma$ on γ, hence $\mathrm{Deg}(\gamma^2)_F = 46$. Moreover, $H^i(\mathcal{O}_F(\gamma))$ is dual to $H^{2-i}(\mathcal{O}_F(-\gamma))$ by Serre duality. This cohomology group can be computed from the exact sequence:

$$0 \rightarrow \mathcal{O}_F(-\gamma) \rightarrow \mathcal{O}_F \rightarrow \mathcal{O}_\gamma \rightarrow 0.$$

It follows that both are zero, hence $\dim|\gamma|_F = 24$. Therefore, indeed, the set of pairs (γ, F) has dimension at most $33 + 24 = 57$.[3]

Now consider case (b). Such a γ must be contained in a reducible quartic, hence in a plane, a quadric, or a cubic surface. The first two possibilities are readily checked and it happens that they contain no curves of the required degree and genus. Moreover, such a curve is contained in a *unique* cubic surface F, because $\mathrm{Deg}(\gamma) = 14 > 9 = \mathrm{Deg}(F' \cdot F'')$, for two distinct cubic surfaces F' and F''. We will say that γ is of type (b_0) if the cubic F is non-singular; otherwise, we will say that γ is of type (b_1).

(C) *Every maximal algebraic family of curves γ of type (b_0) has dimension* 56.

Proof. Let γ be a curve of type (b_0), and let F be the corresponding cubic surface. Since $K_F \equiv -H$, by the Riemann-Roch theorem on F:

$$\dim|\gamma|_F = \frac{\mathrm{Deg}(\gamma \cdot \gamma + \mathrm{H})_F}{2} + \{\dim H^1(\mathcal{O}_F(\gamma)) - \dim H^2(\mathcal{O}_F(\gamma))\}.$$

But $K_\gamma \equiv \gamma \cdot (\gamma + K_F)$, hence $46 = \mathrm{Deg}(\gamma^2)_F - \mathrm{Deg}(\gamma \cdot H)_F = \mathrm{Deg}(\gamma^2)_F - 14$, hence $\mathrm{Deg}(\gamma^2)_F = 60$. Also, $H^i(\mathcal{O}_F(\gamma))$ is dual to $H^{2-i}(\mathcal{O}_F(-H-\gamma))$, and this group can be computed from the exact sequence:

$$0 \rightarrow \mathcal{O}_F(-H-\gamma) \rightarrow \mathcal{O}_F \rightarrow \mathcal{O}_{(H+\gamma)} \rightarrow 0.$$

Since $H + \gamma$ is a reduced and connected curve, $H^0(\mathcal{O}_{H+\gamma}) = k$ (constants), and this implies $H^i(\mathcal{O}_F(-H-\gamma)) = (0)$ for $i = 1$ and 2. Putting all this together, we see that $\dim|\gamma|_F = 37$. Since there is a 19-dimensional family

---

[3] It may be objected that we have used the Riemann-Roch theorem, and Serre duality as though F were non-singular. But since F is non-singular along γ, the former can be proved by means of the exact sequence:

$$0 \rightarrow \mathcal{O}_F \rightarrow \mathcal{O}_F(\gamma) \rightarrow \mathcal{O}_\gamma((\gamma^2)_F) \rightarrow 0.$$

And the latter can be proven either (a) directly by resolving the singularities of F and comparing the cohomology on F and on its resolution, or (b) as a consequence of Grothendieck's general theory [8]. In the second case, one merely has to note that F is always a Cohen-Macauley variety; and since it is a *quartic* surface the canonical sheaf is simply $\mathcal{O}_F$ itself.

of cubic surfaces, (C) follows if we show that a generic γ in a maximal algebraic family is contained in a generic cubic surface. But let $\gamma \subset F$ be any curve of the family. Then recalling that the divisor class group of any non-singular cubic surface is the same as that of any other, it follows that if the set of all non-singular cubic surfaces are suitably parametrized the invertible sheaf $\mathcal{O}_F(\gamma)$ will be a specialization of an invertible sheaf L defined on the generic cubic surface F^*. And since $H^i(\mathcal{O}_F(\gamma)) = (0)$ for $i = 1$ and 2, by the upper semi-continuity of cohomology [9], we conclude that $H^i(L) = (0)$ for $i = 1$ and 2, and that all sections of $\mathcal{O}_F(\gamma)$ are specializations of sections of L. Therefore $\dim H^0(L) = 38$; and since almost all sections of $\mathcal{O}_F(\gamma)$ are non-singular, so are almost sections of L. Hence there is a non-singular $\gamma^* \subset F^*$ specializing to $\gamma \subset F$. QED.

Now suppose C is the Chow variety of non-singular curves of degree 14, and genus 24. Let $C_b \subset C$ be the locus of curves of type (b), and let $C_{b_1} \subset C_b$ be the locus of curves of type (b_1). Then it is clear that C_b and C_{b_1} are closed (possibly reducible) subvarieties of C. By (B) and (C), every component of $C - C_b$ has dimension $\leqq 56$, and every component of $C_b - C_{b_1}$ has dimension $= 56$. Therefore if C_0 equals C minus C_{b_1} and minus the closure of $C - C_b$, C_0 is an open set in the Chow variety, of dimension 56, and parametrizing *almost all* curves of type (b_0).

We shall now single out a set of components of C_0 such that, if N is the normal sheaf to a γ in one of these components, then $\dim H^0(N) = 57$. In fact, we say that $\gamma \subset F$ is of type (b'_0) if there is a line E on F such that $\gamma \equiv 4H + 2E$ on F. Then the corresponding $C'_0 \subset C_0$ which is the locus of such curves is clearly closed in C_0. But it is also open: if $\gamma^* \subset F^*$ specializes to $\gamma \subset F$, and if $\gamma \equiv 4H + 2E$ on F, then first of all, there is a line $E^* \subset F^*$ (possibly only rationally defined after a suitable base extension) which specializes to E; and secondly, since the divisor class group is discrete and constant for all non-singular cubics,

$$\gamma - 4H - 2E \equiv 0 \text{ implies } \gamma^* - 4H^* - 2E^* \equiv 0.$$

Therefore γ^* is of type (b'_0).

(D) *If* $\gamma \subset F$ *is of type* (b'_0), *then* $\dim H^0(N) = 57$.

Proof. Let N_F be the sheaf of normal vector fields to γ and *in* F, and let N_P be the sheaf of normal vector fields to F, and in P_3, which are defined along γ. Then we have the sequence:

$$0 \to N_F \to N \to N_P \to 0.$$

But if D is a non-singular divisor on a non-singular variety V, then its normal sheaf is isomorphic to $\mathcal{O}_D((D^2)_V)$. Therefore $N_F \cong \mathcal{O}_\gamma((\gamma^2)_F)$ and $N_F \cong \mathcal{O}_\gamma(3h)$. But since $K_\gamma \equiv (\gamma^2)_F + \gamma \cdot K_F \equiv (\gamma^2)_F - h$, it follows that $(\gamma^2)_F$ is a non-special divisor, of degree 60 in fact. Therefore $H^1(N_F) = (0)$ and $\dim H^0(N_F) = 60 - (24 - 1) = 37$. On the other hand, by the Riemann-Roch theorem for curves,

$$\begin{aligned} \dim H^0(\mathcal{O}_\gamma(3h)) &= 42 - (24 - 1) + \dim H^0(\mathcal{O}_\gamma(K_\gamma - 3h)) \\ &= 19 + \dim H^0(\mathcal{O}_\gamma((\gamma^2)_F - 4h)) \\ &= 19 + \dim H^0(\mathcal{O}_\gamma(2\gamma \cdot E)) \end{aligned}$$

(using the hypothesis $\gamma \equiv 4H + 2E$). But now, use the exact sequence:

$$0 \to \mathcal{O}_F(-4H) \to \mathcal{O}_F(2E) \to \mathcal{O}_\gamma(2\gamma \cdot E) \to 0.$$

It is readily seen that $H^i(\mathcal{O}_F(-4H)) = (0)$ for $i = 0$ and 1, and that $\dim H^0(\mathcal{O}_F(2E)) = 1$. Putting all this information together we conclude: $\dim H^0(N) = 37 + 19 + 1 = 57$. QED.

It remains only to note:

(E) *If F is any non-singular cubic surface, and $E \subset F$ is any line, there exist non-singular curves $\gamma \in |4H + 2E|$, and they have degree* 14 *and genus* 24.

Proof. The degree and genus of such a γ are computed by the usual formulae, recalling that $\mathrm{Deg}(E^2)_F = -1$. To see that such a γ exists, it suffices, by the characteristic 0 Bertini theorem, to prove that $|4H + 2E|$ has no base points. But the only possible base points are the points of E, and we use the exact sequence:

$$0 \to \mathcal{O}_F(4H + E) \to \mathcal{O}_F(4H + 2E) \to \mathcal{O}_E(2) \to 0.$$

Since the sections of $\mathcal{O}_E(2)$ have no base points, it suffices to prove $H^1(\mathcal{O}_F(4H + E)) = (0)$. But this follows from the sequence:

$$0 \to \mathcal{O}_F(4H) \to \mathcal{O}_F(4H + E) \to \mathcal{O}_E(3) \to 0,$$

since $H^1(\mathcal{O}_F(4H)) = (0)$, and $H^1(\mathcal{O}_E(3)) = (0)$. QED.

HARVARD UNIVERSITY.

REFERENCES.

[1] D. Mumford, "Pathologies of modular algebraic surfaces," *American Journal of Mathematics*, vol. 83 (1961), p. 339.

[2] J. I. Igusa, "Betti and Picard numbers of abstract algebraic surfaces," *Proceedings of the National Academy of Sciences*, vol. 46 (1960), p. 724.

[3] K. Kodaira, "A theorem of completeness of characteristic systems for analytic families of compact submanifolds of complex manifolds," *Annals of Mathematics*, vol. 72 (1962), p. 146.

[4] A. Grothendieck, "Techniques de construction et théorèmes d'existence en géométrie algébrique, IV: Les schémas de Hilbert," *Séminaire Bourbaki*, Exposé 221, 1961.

[5] K. Kodaira, "On stability of compact submanifolds of complex manifolds," to appear in *American Journal of Mathematics*.

[6] M. Noether, *Zur Grundlegung der Theorie der Algebraischen Raumkurven*, Steiner Preisschrift, Berlin, 1882.

[7] S. Lefschetz, "On certain numerical invariants of algebraic varieties with applications to Abelian varieties," (Memoire Bordin), *Transactions of the American Mathematical Society*, vol. 22 (1921), p. 327.

[8] A. Grothendieck, "Théorèmes de dualité pour les faisceaux algebriques cohérents," *Séminaire Bourbaki*, Exposé 149, 1957.

[9] W. L. Chow and J. I. Igusa, "Cohomology theory of varieties over rings," *Proceedings of the National Academy of Sciences*, vol. 44 (1958), p. 1244.

EXPONENTIAL SUMS IN MANY VARIABLES.*

By H. DAVENPORT and D. J. LEWIS.[1]

---

1. In this paper we shall be concerned with sums of the type

$$S(F)=\sum_{x_1}\cdots\sum_{x_n} e^{2\pi i F(x_1\cdots x_n)/p}, \tag{1}$$

where p is a prime and $F(x_1,\cdots,x_n)$ is a polynomial, and each variable of summation runs through a complete set of residues $(\operatorname{mod} p)$. We denote the finite field of p elements by $[p]$ and we suppose that F has coefficients in $[p]$. We suppose further that F has fixed degree k, and that p is large compared with n and k. We do *not* suppose that F itself is fixed independently of p.

Plainly $S(F)$ is invariant under any non-singular linear transformation (homogeneous or not) of $x_1,\cdots,x_n$ with coefficients in $[p]$. We can suppose that F does not degenerate, that is, that F is not expressible by such a transformation as a polynomial in fewer than n variables; for the effect of summation over superfluous variables is merely to give a power of p as a factor.

If $k=2$, so that F is a quadratic polynomial, the value of $S(F)$ is well known. We can write

$$F(x_1,\cdots,x_n)=Q(x_1,\cdots,x_n)+2b_1x_1+\cdots+2b_nx_n, \tag{2}$$

after omitting the constant term, as we may. If Q does not degenerate we find (on expressing Q as a diagonal form) that

$$S(F)=(\epsilon\sqrt{p})^n\left(\frac{\Delta}{p}\right)e^{-2\pi i Q^*(b)/p}, \tag{3}$$

where $\Delta=\det Q$ and Q^* is the adjoint form, and $\epsilon=1$ or i according as $p\equiv 1$ or $-1 \pmod 4$. If Q degenerates and F does not, then $S(F)=0$.

Our main concern is with the case when F is a cubic polynomial, say

$$F=C+Q+L, \tag{4}$$

where C is a cubic form, Q a quadratic form, and L a linear form in $x_1,\cdots,x_n$. Our main object is to establish results of the form

$$|S(F)|\leqq c(n)p^{n-k}, \tag{5}$$

---

* Received July 15, 1962.

[1] An earlier version of this paper was written during a period when the authors received support from a National Science Foundation grant to the University of Notre Dame.

where k is large (if possible) when n is large. In contrast to the quadratic case, some condition (in addition to the non-degeneracy of F) is essential to ensure the validity of such an estimate. If, for example,

$$(6) \qquad F_1 = x_1 G_1(x_{\lambda+1}, \cdots, x_n) + \cdots + x_\lambda G_\lambda(x_{\lambda+1}, \cdots, x_n),$$

where the G_i are quadratic polynomials, then

$$S(F_1) = p^\lambda N,$$

where N denotes the number of solutions of

$$G_1(x_{\lambda+1}, \cdots, x_n) = 0, \cdots, G_\lambda(x_{\lambda+1}, \cdots, x_n) = 0$$

in $[p]$. If λ is small compared with n, then for 'general' quadratic polynomials $G_1, \cdots, G_\lambda$, we have $N \sim p^{n-2\lambda}$ as $p \to \infty$; and it is easy to give particular examples for which this holds, by taking the quadratic parts of the G_i to be suitable diagonal forms. Then

$$(7) \qquad S(F_1) \sim p^{n-\lambda},$$

and so no substantial improvement on the trivial estimate p^n is possible if λ is small. The situation is similar, in principle, for the more general form

$$(8) \qquad F_2 = x_1 G_1(x_1, \cdots, x_n) + \cdots + x_\lambda G_\lambda(x_1, \cdots, x_n).$$

If $x_1, \cdots, x_\lambda$ are all 0, summation over the other variables gives $p^{n-\lambda}$. If $x_1, \cdots, x_\lambda$ are not all 0, summation over the other variables gives $O(p^{(n-\lambda+s)/2})$, where $n - \lambda - s$ is the rank of the part of F_2 which is quadratic in $x_{\lambda+1}, \cdots, x_n$. In general, we can expect that the number of sets of $x_1, \cdots, x_\lambda$ for which this rank is $n - \lambda - s$ will be $O(p^{\lambda-s})$. This gives

$$(9) \qquad S(F_2) = p^{n-\lambda} + O(p^{(n+\lambda)/2}),$$

and again no significant improvement on the trivial estimate is possible if λ is small compared with n.

Our main result is expressed in terms of a certain invariant $h(C)$ of the cubic part $C(\boldsymbol{x})$ of $F(\boldsymbol{x})$. We define $h = h(C)$ to be the least number for which $C(\boldsymbol{x})$ is representable identically as

$$(10) \qquad L_1 Q_1 + \cdots + L_h Q_h,$$

where $L_1, \cdots$ and $Q_1, \cdots$ are linear and quadratic forms respectively, with coefficients in $[p]$. In other words, $n - h$ is the greatest dimension of a linear space on which $C(\boldsymbol{x})$ vanishes identically. Plainly $h(C)$ is an invariant of C, and $0 \leqq h(C) \leqq n$, and $h(C) = 0$ if and only if C vanishes identically, and $h(C) = n$ if and only if C does not represent 0 non-trivially in $[p]$. By Chevalley's theorem ([**3**], 55-57) we have $h < n$ if $n \geqq 4$, and it can be deduced from the work of R. Brauer [**2**] that

$$h \leqq n + 7/2 - \sqrt{(2n + 25/4)}. \tag{11}$$

We prove:

THEOREM 1. *For a cubic polynomial $F(\mathbf{x})$, expressed as in* (4), *we have*[2]

$$|S(F)| \ll p^{n-\frac{1}{4}h}, \tag{12}$$

where $h = h(C)$.

The proof is based on a recent method of Davenport [**5**], adapted for the present purpose. In its original form the method was not applicable to polynomials over $[p]$ for two reasons: (i) it was assumed that the cubic part of F does not represent zero, and (ii) it was assumed that the cubic part of F has fixed integral coefficients, independent of p (see [**5**], end of § 7).

An earlier version of the present paper was based on the less powerful method of [**4**]; and we proved a result similar to (12) but with a 'saving' in the exponent proportional to $\sqrt{h}$ instead of to h.

For $n = 2, 3$ we supplement Theorem 1 by proving the more precise results:

THEOREM 2. *Suppose F does not degenerate. Then for $n = 2$ we have*

$$|S(F)| \ll p^{\frac{5}{4}} \tag{13}$$

and for $n = 3$ we have

$$|S(F)| \ll p^{2}. \tag{14}$$

The result (14) is best possible, if no other condition is imposed on F; this can be seen from the examples (6) or (8) with $\lambda = 1$.

In connection with Theorem 1, it may be remarked as a matter of interest that if an invariant h is defined in the same way for quadratic forms as for cubic forms, then it is easily seen that $\frac{1}{2}n \leqq h \leqq \frac{1}{2}n + 1$ for any non-degenerate form, so that the possibility of h being small when n is large does not arise for quadratic polynomials.

---

[2] We use Vinogradov's notation $\ll$ to indicate an inequality with an unspecified constant factor, which in the present instance depends only on n.

On the other hand, the invariant h (similarly defined) is not relevant to the quartic case, in the sense that h may be large but the sum not small. For example, if

$$F=\Big(\sum_{i=1}^{m}x_ix_{m+i}\Big)^2-\nu\Big(\sum_{i=1}^{m}x_{2m+i}x_{3m+i}\Big)^2,$$

where ν is a non-square, then $h(F)=2m$ and $|S(F)|>>p^{4m-1}$.

It is reasonable to expect that if C degenerates and F does not, then a more precise result than (12) should hold. We prove such a supplementary result in § 5, namely that if C is equivalent to a form in $n_1<n$ variables, the exponent in (12) can be reduced by an amount $\frac{1}{2}(n-n_1)$.

In § 6 we give two immediate applications of Theorem 1:

THEOREM 3. *Let* $F=C+Q+L+c$ (*where* c *is a constant*) *be a cubic polynomial over* $[p]$. *Then the number of solutions of* $F=0$ *in* $[p]$ *is*

$$p^{n-1}+O(p^{n-\frac{1}{4}h}). \tag{15}$$

where $h=h(C)$.

THEOREM 4. *Let* χ *be any non-principal character* (mod p). *Then, with the notation of Theorem* 3,

$$\sum_{x_1,\cdots,x_n}\chi(F(x_1,\cdots,x_n))=O(p^{n-\frac{1}{4}h+\frac{1}{2}}). \tag{16}$$

Another method for the estimation of exponential sums in many variables was developed by Birch [1]. Let F be a polynomial of degree k over $[p]$ and let f be the form of degree k consisting of its terms of highest degree. The equations

$$\partial f/\partial x_1=0,\cdots,\partial f/\partial x_n=0$$

define the singular locus of the cone $f=0$ in n dimensional space over $[p]$. Let s denote the dimension of this singular locus, considered as an algebraic set of points over the algebraic closure of $[p]$. Then Birch's method leads to the estimate

$$|S(F)|<<p^{n-K(n-s)}, \tag{17}$$

where $K=2^{-k+1}$. A detailed proof presents no serious difficulty, but requires some simple auxiliary results on algebraic geometry over $[p]$. The estimate (17), though more general than (12), involves an invariant which is less simply related to the structure of the form than is the invariant $h(C)$ of a cubic form. It is perhaps of interest to note that $\frac{1}{2}(n-s)\leqq h(C)$.

The results of the present paper extend to exponential sums over general finite fields $[q]$, where $q = p^f$. The general character of the additive group of $[q]$ is $e_p(S(\lambda\xi))$, where λ is a fixed element and ξ is a variable element of $[q]$, and where S denotes the trace from $[q]$ to $[p]$ and $e_p(x) = e^{2\pi ix/p}$. Thus the analogous exponential sum is

$$\sum_{\xi_1,\cdots,\xi_n} e_p(S(F(\xi_1,\xi_2,\cdots,\xi_n))), \tag{18}$$

where F is a polynomial with coefficients in $[q]$, and $\xi_1,\xi_2,\cdots,\xi_n$ run through $[q]$. When F is quadratic the sum can be evaluated in much the same way as was indicated for quadratics over $[p]$. When F is cubic, the estimates (with q replacing p) of the present paper apply. The first equation needed for the proof of the extended Theorem 4 can be found in [**6**].

The methods of this paper are essentially elementary and the results may very well be obtainable in easier fashion using more powerful methods.

2. LEMMA 1. *Let* $F = C + Q + L$, *as in* (4), *and let*

$$C(\mathbf{x}) = \sum_i \sum_j \sum_k c_{ijk} x_i x_j x_k, \tag{19}$$

where i, j, k *go from* 1 *to* n *and the coefficients* c_{ijk} *are symmetrical in the* 3 *suffices. Suppose*

$$|S(F)| \geqq Ap^{n-m}. \tag{20}$$

Then there exist at least A^4p^{2n-4m} *pairs of points* $\mathbf{x}$, $\mathbf{y}$ *over* $[p]$ *which satisfy*

$$B_j(\mathbf{x} \mid \mathbf{y}) = \sum_i \sum_k c_{ijk} x_i y_k = 0 \qquad (j = 1,\cdots,n). \tag{21}$$

Proof. We have

$$\begin{aligned} |S(F)|^2 &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} e_p(F(\mathbf{x}+\mathbf{y}) - F(\mathbf{x}-\mathbf{y})) \\ &\leqq \sum_{\mathbf{y}} \Big| \sum_{\mathbf{x}} e_p(\Phi_{\mathbf{y}}(\mathbf{x})) \Big|, \end{aligned}$$

where

$$\Phi_{\mathbf{y}}(\mathbf{x}) = 6 \sum_i \sum_j \sum_k c_{ijk} x_i x_j y_k + 4 \sum_i \sum_j a_{ij} x_i y_j,$$

if

$$Q(\mathbf{x}) = \sum_i \sum_j a_{ij} x_i x_j.$$

By Cauchy's inequality

$$\begin{aligned} |S(F)|^4 &\leqq p^n \sum_{\mathbf{y}} \Big| \sum_{\mathbf{x}} e_p(\Phi_{\mathbf{y}}(\mathbf{x})) \Big|^2 \\ &\leqq p^n \sum_{\mathbf{y}} \sum_{\mathbf{z}} \Big| \sum_{\mathbf{x}} e_p(\Phi_{\mathbf{y},\mathbf{z}}(\mathbf{x})) \Big|, \end{aligned}$$

where

$$\Phi_{\mathbf{y},\mathbf{z}}(\mathbf{x}) = \Phi_{\mathbf{y}}(\mathbf{x}+\mathbf{z}) - \Phi_{\mathbf{y}}(\mathbf{x}-\mathbf{z})$$
$$= 24\sum_i\sum_j\sum_k c_{ijk}x_jz_iy_k + \Psi(\mathbf{y},\mathbf{z}).$$

Summation over x_j gives 0 unless $B_j(\mathbf{y}\mid\mathbf{z}) = 0$, in which case it gives p. Hence

$$|S(F)|^4 \leqq p^n p^n N.$$

where N denotes the number of pairs of points $\mathbf{y}$, $\mathbf{z}$ in $[p]$ with $B_j(\mathbf{y}\mid\mathbf{z}) = 0$ for all j. Hence the result.

The following two lemmas are similar to Lemmas 2 and 3 of [5], but there is enough difference between the two situations to warrant the proofs being given, rather than a mere reference.

LEMMA 2. *Let $f_1(\mathbf{x}),\cdots,f_N(\mathbf{x})$ be forms in $\mathbf{x} = (x_1,\cdots,x_n)$ with coefficients in $[p]$, and suppose that N and the degrees of the forms f_h are bounded in terms of n. Suppose there exist Bp^{n-t} points $\mathbf{x}\neq 0$ in $[p]$ for which*

(22) $$f_1(\mathbf{x}) = 0,\cdots,f_N(\mathbf{x}) = 0,$$

where B is greater than a certain function of n, and t is a positive integer. Suppose also that p is greater than a certain function of n. Then there is one of these points $\mathbf{x}$ for which the rank of the matrix

(23) $$(\partial f_h/\partial x_\nu) \qquad (h = 1,\cdots,N;\nu = 1,\cdots,n)$$

is at most $t-1$.

Proof. The equations (22) define an algebraic set of points in n dimensional space over $[p]$. Such a set is representable as the union of a finite number of absolutely irreducible algebraic varieties, and it follows from the explicit process for determining these ([9], § 31) that, with the present hypotheses, their number is bounded in terms of n. Hence we can choose one of these absolutely irreducible components, say $\mathfrak{V}$, which contains more than $B'p^{n-t}$ points of the given set, where B' is large when B is large.

Let s denote the dimension of $\mathfrak{V}$. Then $\mathfrak{V}$ contains $\ll p^s$ points with coordinates in $[p]$. Hence, if B is greater than a certain function of n, we must have $s \geqq n-t+1$.

Let $\mathfrak{V}$ be given by

$$g_1(\mathbf{x}) = 0,\cdots,g_M(\mathbf{x}) = 0,$$

where the forms $g_1(\mathbf{x}),\cdots,g_M(\mathbf{x})$, with coefficients in some algebraic extension of $[p]$, constitute a basis for the ideal of all polynomials which vanish everywhere on $\mathfrak{B}$. It is well known ([**11**] Ch. 1. Theorem 2) that the rank of the matrix

$$(\partial g_k/\partial x_\nu) \qquad (k=1,\cdots,M;\nu=1,\cdots,n)$$

is at most $n-s\leqq t-1$ at each point of $\mathfrak{B}$. Since each polynomial f_h vanishes everywhere on $\mathfrak{B}$, we have

$$f_h(\boldsymbol{x})=\sum_{k=1}^{M}F_{h,k}(\boldsymbol{x})g_k(\boldsymbol{x})$$

identically, for some polynomials $F_{k,k}$. It follows that the rank of the matrix (23) is at most $t-1$ at each point of $\mathfrak{B}$. The result follows, on taking $\boldsymbol{x}$ to be any one of the points of the given set which is on $\mathfrak{B}$.

LEMMA 3. *Suppose there exist Dp^{n-t} points $\boldsymbol{x}\neq 0$ with coordinates in $[p]$ for which the bilinear equations* (21) *have exactly r linearly independent solutions in $\mathbf{y}$. Then, if D and p are sufficiently large in relation to n, we have*

$$h(C)\leqq n-r+t-1. \tag{24}$$

Proof. For the points $\boldsymbol{x}$ in question, the rank of the matrix

$$\Gamma=(\sum_i c_{ijk}x_i) \qquad (j,k=1,\cdots,n) \tag{25}$$

is $n-r$. We can suppose that there is a set $\mathfrak{X}$ of more than $D'p^{n-t}$ points $\boldsymbol{x}$ for which every subdeterminant of Γ of order $n-r+1$ is 0, and a particular subdeterminant of Γ of order $n-r$ is not 0, where D' is large with D. For any point of $\mathfrak{X}$ we can obtain r linearly independent solutions $\mathbf{y}^{(1)},\cdots,\mathbf{y}^{(r)}$ of (21) by taking the coordinates of these points to be certain particular subdeterminants of Γ of order $n-r$.

For a general point $\boldsymbol{x}$, whose n coordinates are independent indeterminates, if we substitute for $\mathbf{y}$ one of these points, the left hand sides of (21) become certain subdeterminants of Γ of order $n-r+1$. Thus we have, identically in $\boldsymbol{x}$,

$$\sum_i\sum_k c_{ijk}x_iy_k^{(p)}=\Delta_{j,p}(\mathbf{x}) \tag{26}$$

for $j=1,\cdots,n$ and $p=1,\cdots,r$, where the $\Delta_{j,p}$ are certain subdeterminants of (25) of order $n-r+1$ (allowing repetitions of rows).

Since $x_1,\cdots,x_n$ are independent variables in (26), partial differentiation with respect to x_ν gives

$$\sum_k c_{\nu jk}y_k^{(p)}+\sum_i\sum_k c_{ijk}x_i\frac{\partial y_k^{(p)}}{\partial x_\nu}=\frac{\partial}{\partial x_\nu}\Delta_{j,p},$$

for $j=1,\cdots,n$ and $\nu=1,\cdots,n$ and $p=1,\cdots,r$. Multiplying by $y_j^{(q)}$ and summing over j, and noting that

$$\sum_i \sum_j c_{ijk} x_i y_j^{(q)} = \Delta_{k,q}$$

by (26), we obtain

$$\sum_j \sum_k c_{\nu jk} y_k^{(p)} y_j^{(q)} + \sum_k \Delta_{k,q} \frac{\partial y_k^{(p)}}{\partial x_\nu} = \sum_j y_j^{(q)} \frac{\partial}{\partial x_\nu} \Delta_{j,p} \tag{27}$$

for $\nu=1,\cdots,n$ and $p=1,\cdots,r$ and $q=1,\cdots,r$.

We now appeal to Lemma 2, taking the forms $f_1,\cdots,f_N$ to be all the subdeterminants $\Delta_{j,p}(x)$, for $j=1,\cdots,n$ and $p=1,\cdots,r$. By that lemma, if D is sufficiently large, there is a point x of the set $\mathfrak{X}$ for which all the $\Delta_{j,p}$ vanish and for which the rank of the matrix

$$\left(\frac{\partial}{\partial x_\nu} \Delta_{j,p}\right)$$

is at most $t-1$. The last statement implies that

$$\frac{\partial}{\partial x_\nu} \Delta_{j,p} = \sum_{\tau=1}^{t-1} W_{j,p,\tau} U_{\tau,\nu}$$

for certain sets of numbers W, U. Since the values of the derivatives lie in $[p]$, we can suppose that all the numbers W, U are in $[p]$.

From now onwards we specialize x to be such a point. Then (27) takes the form

$$\sum_j \sum_k c_{\nu jk} y_k^{(p)} y_j^{(q)} = \sum_j y_j^{(q)} \sum_{\tau=1}^{t-1} W_{j,p,\tau} U_{\tau,\nu}.$$

Let $T_1,\cdots,T_r$ be any indeterminates, and put

$$Y = T_1 y^{(1)} + \cdots + T_r y^{(r)}. \tag{28}$$

Multiplying the previous equation by $T_p T_q$ and summing over p and q, we obtain

$$\begin{aligned} \sum_j \sum_k c_{\nu jk} Y_j Y_k &= \sum_j \sum_{p=1}^{r} \sum_{q=1}^{r} T_p T_q y_j^{(q)} \sum_{\tau=1}^{t-1} W_{j,p,\tau} U_{\tau,\nu} \\ &= \sum_{\tau=1}^{t-1} V_\tau U_{\tau,\nu}, \end{aligned}$$

say, and this holds for $\nu=1,\cdots,n$. Multiplying by Y_ν and summing over ν, and using (28) on the right, we obtain

$$\sum_\nu \sum_j \sum_k c_{\nu jk} Y_\nu Y_j Y_k = \sum_{\tau=1}^{t-1} V_\tau \sum_{p=1}^{r} T_p \sum_{\nu=1}^{n} y_\nu^{(p)} U_{\tau,\nu}.$$

If the indeterminates $T_1, \cdots, T_r$ satisfy the $t-1$ homogeneous linear equations

$$\sum_{\rho=1}^{r} T_\rho \sum_{\nu=1}^{n} y_\nu^{(\rho)} U_{\tau,\nu} = 0 \quad (\tau = 1, \cdots, t-1),$$

then $C(\mathbf{Y}) = 0$. Referring to (28), we see that there is a linear space for $\mathbf{Y}$ of dimension at least $r-(t-1)$ on which $C(\mathbf{Y})$ vanishes. It follows from the definition of $h(C)$ that

$$n - h(C) \geqq r - t + 1,$$

which proves (24).

3. *Proof of Theorem* 1. Suppose (20) holds, where $4m$ is an integer not exceeding n. Then there exist at least $(A^4-1)p^{2n-4m}$ pairs $\boldsymbol{x}, \boldsymbol{y}$ of points with coordinates in $[p]$ which satisfy (21), and for which $\boldsymbol{x} \neq 0$.

We divide these pairs into classes, according to the number of linearly independent solutions of (21) for given $\boldsymbol{x}$. If this number is r $(1 \leqq r \leqq n)$, then there are exactly p^r points $\boldsymbol{y}$ associated with a given point $\boldsymbol{x}$ as solutions of (21). Hence, for some $\boldsymbol{x} \neq 0$ and some r there are at least $n^{-1}(A^4-1)p^{2n-4m-r}$ points $\boldsymbol{x}$ in the class r.

Thus the hypotheses of Lemma 3 are satisfied, with $t = 4m + r - n$. It follows from (24), provided A is sufficiently large, that

$$h(C) \leqq n - r + (4m + r - n) - 1 = 4m - 1.$$

We therefore reach a contradiction if (20) holds with $m = \frac{1}{4}h$. (The condition $4m \leqq n$ is automatically satisfied.) This proves Theorem 1.

4. *Proof of Theorem* 3. The number of solutions of $F = 0$ in $[p]$ is

$$\begin{aligned}\sum_{x_1, \cdots, x_n} p^{-1} \sum_u e_p(uF(x_1, \cdots, x_n)) &= p^{n-1} + p^{-1} \sum_{u \neq 0} S(uF) \\ &= p^{n-1} + O(p^{n-\frac{1}{4}h}),\end{aligned}$$

where $h = h(C)$.

Proof of Theorem 4. It is well known ([7]), 486, 492) that

$$\chi(z) = \frac{1}{\tau(\bar{\chi})} \sum_u \bar{\chi}(u) e(uz)$$

where

$$|\tau(\bar{\chi})| = p^{\frac{1}{2}}.$$

Hence

$$\begin{aligned}\sum_{x_1, \cdots, x_n} \chi(F(x_1, \cdots, x_n)) &= \frac{1}{\tau(\bar{\chi})} \sum_u \bar{\chi}(u) S(uF) \\ &= O(p^{n-\frac{1}{4}h+\frac{1}{2}}).\end{aligned}$$

As a corollary, we may note that, if t is any fixed divisor of $p-1$, the number of solutions of

$$y^t \equiv F(x_1, \cdots, x_n) \pmod{p} \tag{29}$$

is

$$p^n + O(p^{n-\frac{1}{4}h+\frac{1}{2}}). \tag{30}$$

5. We now prove (as Theorem 5 below) the supplementary result concerning sums for which $C(\boldsymbol{x})$ degenerates but $F(\boldsymbol{x})$ does not. It is convenient to prove first a simple property of the invariant $h(C)$.

Lemma 4. (a) *Let $C(x_1, \cdots, x_n)$ be a cubic form, expressible as $C_1(y_1, \cdots, y_{n_1})$ by a non-singular linear transformation. Then*

$$h(C) = h(C_1). \tag{31}$$

(b) *For $\nu < n$, let*

$$C^*(x_1, \cdots, x_\nu) = C(x_1, \cdots, x_\nu, 0, \cdots, 0). \tag{32}$$

Then

$$h(C) - (n-\nu) \leqq h(C^*) \leqq h(C). \tag{33}$$

Proof. Any representation of C_1 in the form (10) implies a corresponding representation of C with the same h, whence $h(C) \leqq h(C_1)$. On the other hand, a representation of C in the form (10) implies a representation of C_1, in which $Q_1, \cdots, L_1, \cdots$ are quadratic and linear forms in $y_1, \cdots, y_n$. Since this is an identity, we can put $y_{n_1+1} = \cdots = y_n = 0$ and get a representation of $C_1(y_1, \cdots, y_{n_1})$. Hence $h(C_1) \leqq h(C)$.

As regards (b), it is obvious that any representation of C implies a representation of C^*, whence $h(C^*) \leqq h(C)$. Also we can obviously express $C(x_1, \cdots, x_n)$ as

$$C^*(x_1, \cdots, x_\nu) + x_{\nu+1}Q_{\nu+1}(x_1, \cdots, x_n) + \cdots + x_nQ_n(x_1, \cdots, x_n),$$

whence $h(C) \leqq h(C^*) + n - \nu$.

Theorem 5. *Let $F(x_1, \cdots, x_n)$ be a cubic polynomial over $[p]$, expressed as in (4), which is non-degenerate, and suppose that $C(x_1, \cdots, x_n)$ is equivalent to a form in $n_1 < n$ variables. Then*

$$|S(F)| \ll p^{n-\frac{1}{2}(n-n_1)-\frac{1}{4}h}, \tag{34}$$

where $h = h(C)$.

Proof. We can take $C(\mathbf{x}) = C(x_1, \cdots, x_{n_1})$. Suppose the quadratic form

$$Q(0, \cdots, 0, x_{n_1+1}, \cdots, x_n)$$

has rank $s - n_1$, where $n_1 \leqq s \leqq n$. Then, after a linear substitution on $x_{n_1+1}, \cdots, x_n$ only, we can express this form as

$$a_{n_1+1}x_{n_1+1}^2 + \cdots + a_s x_s^2,$$

where each a_j is in $[p]$ and not 0. Now

$$Q(x_1, \cdots, x_n) = Q^*(x_1, \cdots, x_{n_1}) + a_{n_1+1}x_{n_1+1}^2 + \cdots + a_s x_s^2 + B(x_1, \cdots, x_{n_1} \mid x_{n_1+1}, \cdots, x_n),$$

where B denotes a bilinear form in the two sets of variables indicated.

We can suppose that $x_{n_1+1}, \cdots, x_s$ do not occur in B, for if e. g. $x_1 x_{n_1+1}$ occurs, it can be removed by a substitution on x_{n_1+1} of the type

$$x_{n_1+1} = x'_{n_1+1} + cx_1.$$

(This changes Q^*, but that is immaterial.) Hence we can suppose that

$$B = x_{s+1}L_{s+1} + \cdots + x_n L_n,$$

where $L_{s+1}, \cdots, L_n$ are linear forms in $x_1, \cdots, x_{n_1}$. We can suppose that these forms are linearly independent (if $s < n$); for if not we could change L_n into 0 by a substitution on $x_{s+1}, \cdots, x_n$, and then the variable x_n would not occur in C or Q. It would therefore have to occur with a non-zero coefficient in L, and summation over x_n would give 0. Since $L_{s+1}, \cdots, L_n$ are linearly independent forms, we can take them to be $x_1, \cdots, x_{n-s}$, by a substitution on $x_1, \cdots, x_{n_1}$. We note that $n - s \leqq n_1$ necessarily, since there cannot be more than n_1 linearly independent linear forms in n_1 variables. We now have

$$Q(x_1, \cdots, x_n) = Q^*(x_1, \cdots, x_{n_1}) + a_{n_1+1}x_{n_1+1}^2 + \cdots + a_s x_s^2 + x_{s+1}x_1 + \cdots + x_n x_{n-s}.$$

Now turn to the linear part $L(x_1, \cdots, x_n)$ of F. Any term in $x_{n_1+1}, \cdots, x_s$ can be removed by the process of 'completing the square' applied to any one of these variables. So we can take

$$L(x_1, \cdots, x_n) = L^*(x_1, \cdots, x_{n_1}) - b_1 x_{s+1} - \cdots - b_{n-s}x_n.$$

Collecting the parts, we have

$$F = F^* + a_{n_1+1}x_{n_1+1}^2 + \cdots + a_s x_s^2 + x_{s+1}(x_1 - b_1) + \cdots + x_n(x_{n-s} - b_{n-s}),$$

where F^* is a cubic polynomial in $x_1, \cdots, x_{n_1}$ whose cubic part is $C(x_1, \cdots, x_{n_1})$.

Summation over $x_{n_1+1}, \cdots, x_s$ gives a factor of absolute value $p^{\frac{1}{2}(s-n_1)}$ in $S(F)$. Summation over $x_{s+1}, \cdots, x_n$ gives 0 unless $x_1 = b_1, \cdots, x_{n-s} = b_{n-s}$, in which case it gives p^{n-s}. Hence

$$(35) \qquad |S(F)| \leqq p^{n-\frac{1}{2}n_1-\frac{1}{2}s} \Big| \sum_{x_{n-s+1}, \cdots, x_{n_1}} e_p(F^*(b_1, \cdots, b_{n-s}, x_{n-s+1}, \cdots, x_{n_1})) \Big| .$$

If $s \leqq n - h$, where $h = h(C)$, we use the trivial estimate $p^{n_1-(n-s)}$ for the sum on the right, and obtain

$$|S(F)| \leqq p^{\frac{1}{2}s+\frac{1}{2}n_1} \leqq p^{n-\frac{1}{2}(n-n_1)-\frac{1}{2}h},$$

which is stronger than (34).

Now suppose that $s > n - h$. The cubic part of

$$F^*(b_1, \cdots, b_{n-s}, x_{n-s+1}, \cdots, x_{n_1})$$

is derived from $C(x_1, \cdots, x_{n_1})$ by replacing $n - s$ of the variables by 0. Hence by Lemma 4(b) its invariant h^* satisfies

$$h^* \geqq h - (n - s).$$

By Theorem 1, the sum on the right of (35) has absolute value

$$\ll p^{n_1-(n-s)-\frac{1}{4}(h-n+s)}.$$

Hence

$$\begin{aligned} |S(F)| &\ll p^{n-\frac{1}{2}n_1-\frac{1}{2}s+n_1-n+s-\frac{1}{4}(h-n+s)} \\ &= p^{\frac{1}{4}n+\frac{1}{2}n_1+\frac{1}{4}s-\frac{1}{4}h} \\ &\leqq p^{n-\frac{1}{2}(n-n_1)-\frac{1}{4}h}, \end{aligned}$$

which is (34). This proves Theorem 5.

6. *Proof of Theorem* 2 $(n = 2)$. We can suppose that the cubic part $C(x_1, x_2)$ of $F(x_1, x_2)$ does not degenerate; for if it does, Theorem 5 is applicable with $n = 2$, $n_1 = 1$, $h = 1$, and (34) gives the desired estimate (13). (We exclude throughout the possibility that the cubic part of F vanishes identically.)

We write

$$\begin{aligned} C(x_1, x_2) &= ax_1^3 + 3bx_1^2x_2 + 3cx_1x_2^2 + dx_2^3, \\ Q(x_1, x_2) &= Ax_1^2 + 2Bx_1x_2 + Cx_2^2, \\ L(x_1, x_2) &= Dx_1 + Ex_2. \end{aligned}$$

We have

$$(36) \qquad |S(F)|^2 = \sum_{y_1, y_2} \sum_{x_1, x_2} e_p(\Psi_{y_1, y_2}(x_1, x_2)),$$

where

$$\begin{aligned}\Psi_{y_1,y_2}(x_1,x_2) &= F(x_1+y_1,x_2+y_2)-F(x_1-y_1,x_2-y_2)\\ &= 6\{(ay_1+by_2)x_1^2+2(by_1+cy_2)x_1x_2+(cy_1+dy_2)x_2^2\}\\ &\quad+4\{(Ay_1+By_2)x_1+(By_1+Cy_2)x_2\}\\ &\quad+2\{ay_1^3+3by_1^2y_2+3cy_1y_2^2+dy_2^3\}\\ &\quad+2\{Dy_1+Ey_2\}.\end{aligned}$$

The binary quadratic form in x_1, x_2 on the second line of the last expression does not degenerate identically in y_1, y_2, for that would imply that

$$\Delta(y_1,y_2) = -(by_1+cy^2)^2+(ay_1+by_2)(cy_1+dy_2)$$

vanished identically, which in turn would imply the degeneration of the binary cubic $C(x_1,x_2)$. Also there is no pair $y_1, y_2 \neq 0, 0$ for which ay_1+by_2, by_1+cy_2, cy_1+dy_2 all vanish, for the same reason.

In the sum on the right of (36), the terms with $y_1=y_2=0$ contribute $O(p^2)$. There may be $O(p)$ pairs y_1, y_2 for which the quadratic part of Ψ in x_1, x_2 has rank 1, and their contribution is

$$O(p\cdot p^{\frac{3}{2}}) = O(p^{\frac{5}{2}}).$$

For all other pairs we have

$$\Delta(y_1,y_2)\neq 0.$$

Carrying out the summation over x_1, x_2 in accordance with (3) we obtain

$$|S(F)|^2 = O(p^{\frac{5}{2}}) + (\epsilon\sqrt{p})^2 \sum_{y_1,y_2}\left(\frac{\Delta(y_1,y_2)}{p}\right) e_p(G(y_1,y_2)), \tag{37}$$

where

$$\begin{aligned}G(y_1,y_2) &= -\tfrac{2}{3}\frac{H(y_1,y_2)}{\Delta(y_1,y_2)}+2(ay_1^3+\cdots)+2(Dy_1+Ey_2),\\ H(y_1,y_2) &= (cy_1+dy_2)(Ay_1+By_2)^2\\ &\qquad -2(by_1+cy_2)(Ay_1+By_2)(By_1+Cy_2)\\ &\qquad\qquad +(ay_1+by_2)(By_1+Cy_2)^2.\end{aligned}$$

The sum over y_1, y_2 in (37) is subject to $\Delta(y_1,y_2)\neq 0$.

The terms with $y_2=0$ contribute $O(p^2)$. In the terms with $y_2\neq 0$, we put $y_1=uy_2$ and afterwards omit the suffix on y_2. The double sum on the right of (37) becomes

$$\sum_u\left(\frac{\Delta(u,1)}{p}\right)\sum_y e_p\left(-\frac{2H(u,1)}{3\Delta(u,1)}\,y+2(au^3+\cdots)y^3+2(Du+E)y\right),$$

subject to $\Delta(u,1)\neq 0$ and $y\neq 0$. There may be $O(1)$ values of u for which $au^3+\cdots+d=0$; for all other u the inner sum over y has absolute value less than $2p^{\frac{1}{2}}$, by a theorem[8] of A. Weil [**10**]. Hence the double sum is $O(p^{\frac{3}{2}})$, and this gives

$$|S(F)|^2 = O(p^{\frac{5}{2}}),$$

whence (13).

Proof of Theorem 2 ($n = 3$). If the cubic part $C(x_1, x_2, x_3)$ of $F(x_1, x_2, x_3)$ degenerates into a form in one variable, we can apply (34) with $n = 3$, $n_1 = 1$, $h = 1$, getting $|S(F)| \ll p^{3-1-\frac{1}{4}}$, which is more precise than (14). If $C(x_1, x_2, x_3)$ degenerates into a form in 2 variables with $h = 2$, then (34) gives $|S(F)| \ll p^{3-\frac{1}{2}-\frac{1}{2}}$, which is (14). There remains the possibility that C degenerates into a form in 2 variables with $h = 1$, which we must argue directly. In this case,

$$\begin{aligned} F = x_1(ax_1^2 + bx_1x_2 + cx_2^2) \\ + a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + a_{33}x_3^2 \\ + b_1x_1 + b_2x_2 + b_3x_3. \end{aligned}$$

Summation over x_2, x_3 gives $|S(F)| \ll p(p^{\frac{1}{2}})^2$, as desired, allowing for $O(1)$ values of x_1 for which

$$a_{23}^2 - (cx_1 + a_{22})a_{33} = 0.$$

This fails only if $a_{23} = a_{33} = 0$ or $c = a_{23}^2 - a_{22}a_{33} = 0$, and the result is easily verified in these special cases. In the former case, summation over x_3 gives 0 except for $O(1)$ values of x_1, and in the latter case, summation over x_2 gives 0 except for $O(1)$ values of x_1.

We can now suppose that $C(x_1, x_2, x_3)$ does not degenerate. This implies, by a classical result (see [**7**]) that the Hessian of the form, namely

$$\Delta(x_1, x_2, x_3) = \det(\sum_i c_{ijk}x_i) \quad (j, k = 1, \cdots, n) \tag{38}$$

does not vanish identically. It implies further that there is no point $\mathbf{x} \neq 0$ for which this determinant is of rank 0. For this would imply that the bilinear equations (21) are satisfied for this $\mathbf{x}$ and any $\mathbf{y}$, and this in turn would imply that $\Delta(y_1, y_2, y_3)$ is identically zero.

As in the preceding proof, we have

$$|S(F)|^2 = \sum_{\mathbf{y}} \sum_{\mathbf{x}} e_p(\Psi_{\mathbf{y}}(\mathbf{x})), \tag{39}$$

where

$$\begin{aligned} \Psi_y(x) = F(x_1 + y_1, x_2 + y_2, x_3 + y_3) - F(x_1 - y, x_2 - y_2, x_3 - y_3) \\ = 6 \sum_i \sum_j \sum_k c_{ijk}y_ix_jx_k \\ + 4 \sum_i \sum_j a_{ij}y_ix_j \\ + 2 \sum_i \sum_j \sum_k c_{ijk}y_iy_jy_k + 2 \sum_i d_iy_i. \end{aligned}$$

---

[8] An appeal to this deep result could be avoided, at the cost of complicating the argument.

We classify the points $\mathbf{y}$ in the sum on the right of (39) according to the rank of the quadratic part in $\mathbf{x}$ of $\Psi_{\mathbf{y}}(\mathbf{x})$, i.e. according to the rank of the determinant $\Delta(\mathbf{y})$. The only point $\mathbf{y}$ for which the rank is 0 is $\mathbf{y}=0$, and the contribution of this to the sum is obviously p^3. We shall prove in Lemma 5 that there are only $O(p)$ points for which the rank is 1, and this will imply that the contribution of such points in $O(p^4)$. There are only $O(p^2)$ points $\mathbf{y}$ for which rank is 2, since $\Delta(y)$ does not vanish identically. For such points, the sum over x is $O(p\cdot p)$, giving a contribution from all such points of $O(p^4)$.

It remains now to consider points $\mathbf{y}$ for which $\Delta(\mathbf{y})\neq 0$. Summing over x, we obtain

$$(40)\qquad (\epsilon\sqrt{p})^3\sum_{\mathbf{y}}\left(\frac{\Delta(\mathbf{y})}{p}\right)e_p(G(\mathbf{y})),$$

where

$$G(\mathbf{y})=-\tfrac{2}{3}\frac{H(\mathbf{y})}{\Delta(\mathbf{y})}+2\sum\sum\sum c_{ijk}y_iy_jy_k+2\sum d_iy_i,$$

$$H(\mathbf{y})=(L_{22}L_{33}-L_{23}{}^2)M_1{}^2+\cdots$$

on writing $L_{jk}=L_{jk}(\mathbf{y})$ for the general element in the determinant (38), with $\mathbf{y}$ in place of x, and

$$M_j(\mathbf{y})=\sum_i a_{ij}y_i.$$

The terms with $y_3=0$ contribute $O(p^{7/2})$ to the sum (40), so we can limit ourselves to terms with $y_3\neq 0$ (and of course $\Delta(\mathbf{y})\neq 0$). Putting $y_1=u_1y_3$, $y_2=u_2y_3$, and then omitting the suffix on y_3, the sum becomes

$$(\epsilon\sqrt{p})^3\sum_{u_1,u_2}\left(\frac{\Delta(u_1,u_2,1)}{p}\right)\sum_{y}\left(\frac{y}{p}\right)e_p(P(u_1,u_2)y^3+Q(u_1,u_2)y),$$

where P, Q are rational functions of u_1, u_2. The inner sum is $O(p^{\frac{1}{2}})$ by the result of A. Weil already quoted (and again this quotation could be avoided if desired). Hence the sum is $O(p^4)$, giving the desired result (14).

It remains to prove:

LEMMA 5. *For a non-degenerate cubic form $C(x_1,x_2,x_3)$, there are $\ll p$ points $\mathbf{x}$ for which the determinant* (38) *has rank* 1.

Proof. Suppose there are more than Ap points $\mathbf{x}$ with the property, where A is a suitable absolute constant. By Lemma 3 with $n=3$, $r=2$, $t=2$ we have $h(C)\leq 2$. Hence C represents 0 for some $\mathbf{x}\neq 0$ in $[p]$.

Without loss of generality we can suppose that $C(1,0,0)=0$. If the highest terms in x_1 with non-zero coefficient are quadratic, we can transform linearly so that they reduce to $x_1^2x_2$. By a further obvious transformation we can remove any terms in $x_1x_2^2$ and $x_1x_2x_3$. Thus we have a form of the type

$$\text{(A)} \qquad 3x_1^2x_2 + 3ax_1x_3^2 + C_1(x_2, x_3).$$

If the highest terms in x_1 are linear, we can transform the cubic into one of the type

$$\text{(B)} \qquad 3x_1(x_2^2 + ax_3^2) + C_1(x_2, x_3).$$

Type (*A*). We take $a = 1$, since the case $a = 0$ is more easily settled. The matrix of the determinant (38) is now

$$\begin{pmatrix} y_2 & y_1 & y_3 \\ y_1 & c_{222}y_2 + c_{223}y_3 & c_{223}y_2 + c_{233}y_3 \\ y_3 & c_{223}y_2 + c_{233}y_3 & y_1 + c_{233}y_2 + c_{333}y_3 \end{pmatrix}.$$

For this to have rank 1, we must have

$$\text{(41)} \qquad y_1^2 - y_2(c_{222}y_2 + c_{223}y_3) = 0,$$

$$\text{(42)} \qquad y_1y_3 - y_2(c_{223}y_2 + c_{233}y_3) = 0,$$

$$\text{(43)} \qquad y_2(y_1 + c_{233}y_2 + c_{333}y_3) - y_3^2 = 0.$$

Eliminating y_1 from (41), (42) we get

$$\text{(44)} \qquad y_2y_3^2(c_{222}y_2 + c_{223}y_3) - y_2^2(c_{223}y_2 + c_{233}y_3)^2 = 0.$$

If $y_2 = 0$ then $y_1 = 0$ by (41), and there are only p possibilities for y_3. If $y_2 \neq 0$, (44) is a cubic equation for y_3/y_2; unless it is identically satisfied there are only $O(p)$ possibilities for y_2, y_3 and then (41) determines y_1. Hence (44) must be an identity, whence

$$c_{223} = 0, \qquad c_{222} = c_{233}^2.$$

Now (41), (42) become

$$y_1^2 - c_{233}^2y_2^2 = 0, \qquad y_1y_3 - c_{233}y_2y_3 = 0.$$

We can suppose $y_3 \neq 0$, since if $y_3 = 0$ there are only $O(p)$ possibilities for y_1, y_2. Hence

$$y_1 = c_{233}y_2.$$

Now (43) becomes

$$y_2(2c_{233}y_2 + c_{333}y_3) - y_3^2 = 0.$$

This determines y_3 in terms of y_2, and then (41) determines y_1. Thus there are only $O(p)$ possibilities for y_1, y_2, y_3.

Type (B). The matrix is now

$$\begin{pmatrix} 0 & y_2 & ay_3 \\ y_2 & y_1 + c_{222}y_2 + c_{223}y_3 & c_{223}y_2 + c_{233}y_3 \\ ay_3 & c_{223}y_2 + c_{233}y_3 & ay_1 + c_{233}y_2 + c_{333}y_3 \end{pmatrix}.$$

For this to have rank 1 we must have

$$y_2 = 0, \qquad ay_3 = 0.$$

So we need only consider the case $a = 0$. We need also

$$(y_1 + c_{223}y_3)(c_{333}y_3) - c_{233}^2 y_3^2 = 0.$$

This gives $O(p)$ possibilities for y_1, y_3 unless it is an identity, in which case

$$c_{333} = 0, \qquad c_{233} = 0.$$

But then the form (B) becomes

$$x_2^2(3x_1 + c_{222}x_2 + 3c_{223}x_3),$$

which obviously degenerates, contrary to hypothesis.

Trinity College, Cambridge;
University of Michigan.

REFERENCES.

[1] B. J. Birch, "Forms in many variables," *Proceedings of the Royal Society*, A, vol. 265 (1962), pp. 245-263.

[2] R. Brauer, "A note on systems of homogeneous algebraic equations," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 749-755.

[3] H. Davenport, *The Higher Arithmetic*. London: Hutchinsons, 1952.

[4] ———, "Cubic forms in 32 variables," *Philosophical Transactions of the Royal Society*, A, vol. 251 (1959), pp. 193-232.

[5] ———, "Cubic forms in 16 variables," submitted to *Proceedings of the Royal Society*, A.

[6] H. Davenport and H. Hasse, "Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen," *Journal für die reine und angewandte Mathematik*, vol. 172 (1934), pp. 151-182.

[7] P. Gordan and M. Nöther, "Ueber die algebraischen Formen, deren Hesse'sche Determinante identisch verschwindet," *Mathematische Annalen*, vol. 10 (1876), pp. 547-568.

[8] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*. Berlin: Teubner, 1909.

[9] B. L. Van der Waerden, *Einführung in die algebraische Geometrie*. Berlin: Springer, 1939.

[10] A. Weil, "On some exponential sums," *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 34 (1944), pp. 204-207.

[11] ———, *Foundations of algebraic geometry*. New York: American Mathematical Society Colloquium Publications, 1946.

ANALYTICITY AND PARTIAL DIFFERENTIAL EQUATIONS, I.*

By Felix E. Browder.[1]

Introduction. In the present paper we consider a family of partial differential operators $\{A_t\}$ in a vector bundle B over a manifold M, with the family indexed by and analytic in a parameter t running over a real-analytic manifold M_1. It is our object to present a general method of constructing fundamental solutions $e_t(x,y)$ for A_t and more generally solutions of equations of the form $A_t u_t(x) = f_t(x)$ such that e_t and u_t are analytic in t. The questions which we treat here were suggested obliquely by the paper of Kodaira and Spencer [13] concerning solutions of a differentiable family of elliptic differential operators in a vector bundle over a compact manifold (with applications to variation of complex structure). Our essential results are for the drastically different case of non-compact manifolds and relate to a body of ideas first systematically treated in the Thesis of Malgrange [17] and developed further by the present writer in [3]. Among the results of [17], it was shown that if A is an elliptic differential operator with analytic coefficients in an analytic vector bundle B over a non-compact manifold M, then A has a bi-regular fundamental solution in the large and the equation $Au = f$ has an analytic solution u in the large for every analytic section f. Our concern here is whether such results can be obtained analytically in t if the operators depend analytically on the parameter t.

The specific topics which we discuss are the following:

(I) Does there exist a family of fundamental solutions e_t for A_t in B which depend analytically upon t in M_1 and which for a given parameter value t_0 in M_1, coincides with a prescribed fundamental solution e_0 for A_{t_0}?

(II) If f is an analytic function from M_1 into the analytic sections of B, does there exists a family of solutions $\{u_t\}$ of the equations $A_t u_t = f_t$, with u_t depending analytically upon t in M_1 and such that for a given t_0 in M_1, u_{t_0} coincides with a prescribed solution u_0 of the equation $A_{t_0} u_0 = f_{t_0}$.

(III) If M' is an open subset of M and if $\{v_t\}$ is an analytic family of solutions of the equations $A_t v_t = 0$ over M', can the v_t be approximated

---

* Received September 13, 1962.

[1] The preparation of this paper was partially supported by N.S.F. Grant 19751.

over compact subsets of M' by solutions u_t of $A_t u_t = 0$ over M, with the family $\{u_t\}$ also analytic in t on M_1?

In the present paper, we shall establish an affirmative answer to all three of the above questions for A_t all elliptic and the case when M is a pre-compact open subset of a larger analytic manifold over which all the $\{A_t\}$ have analytic coefficients. We shall treat the case of a general non-compact manifold M in a later paper. Let us remark that the topic (III) above is the analytic generalization of the generalized Runge theorem established for elliptic operators by Lax [14] and Malgrange [17]. In the final section of the present paper, we present a generalization of our results to various broad classes of non-elliptic operators, including operators with constant coefficients, formally hypoelliptic operators, hyperbolic operators, and various classes of non-elliptic, non-hyperbolic operators for which uniqueness of the Cauchy problem and hypoellipticity of the minimal operator have been established.

An important feature of our results which should be emphasized at this point is that the parameter manifold M_1 is given in advance and the analytic families of solutions constructed are analytic in the whole of M_1, not merely upon some arbitrarily small neighborhood of a given point of M_1. The local existence of an analytic family of solutions of the various types considered is obtained fairly obviously by perturbation methods. The existence in the large demands a different treatment.

To illustrate the character of the results which we obtain for families $\{A_t\}$ on pre-compact but non-compact manifolds, it is illuminating to consider one basic type of result established for elliptic operators in Section 5 and for some classes of non-elliptic operators in Section 6, namely: *If for each t in M_1, S_t denotes the closed subspace of $L^2(B)$ consisting of all solutions of $A_t u = 0$, and P_t denotes the orthogonal projection operator of $L^2(B)$ on S_t, then P_t is analytic in t on M_1 and for t near t_0, P_t maps S_{t_0} one-to-one onto S_t.* In the case of a compact manifold M, on the other hand, we might choose $A_t = A - tI$, with t a complex number and A a fixed self-adjoint elliptic operator in $L^2(B)$. Then $S_t = \{0\}$ except for an infinite discrete set of points t, the eigenvalues of A.

The methods used in the present paper are a development of the arguments given by the writer in [3] and, unlike the proofs of Malgrange in [17], make no use of the theory of topological tensor products.

In Section 1, we present the abstract results on analytic famiiles of closed operators in Hilbert space which are the basis of our results on differential operators. Section 2 presents the basic definitions for differential operators and specifically elliptic differential operators in vector bundles over a differentiable manifold. Section 3 gives a brief treatment of elliptic operators on compact manifolds, deriving for this case by our simpler methods the analytic analogues of the theorems of Kodaira and Spencer [13] for the differentiable case. (*Added in proof*: It has been pointed out to the writer by M. Kuranishi that the results of Section 3 answer the questions on analytic dependence raised and left open by A. Weil in his paper "On discrete subgroups of Lie groups (II)," *Annals of Mathematics,* vol. 75, No. 3, May 1962, pp. 578-602 (cf. esp. Appendix III, pp. 598-602 extracted from a letter of Hörmander).) Section 4 presents briefly the proofs of some results on the interrelations of the analyticity in a parameter of realizations of elliptic operators in L^2 and the corresponding analyticity in various senses of the kernels of their resolvents. In Section 5 we establish our principal results for elliptic operators. In Section 6, we give an extension of these results to various classes of non-elliptic operators.

The results of the present paper have been announced by the writer in [5].

1. Construction of analytic families of operators.

Let E be a Banach space, E^* its conjugate space, (e, e^*) the pairing between an element e of E and the functional e^* of E^*. We shall assume E to be a complex Banach space and E^* to be the space of conjugate-linear complex functionals on E, so that $(e, e^*) = \overline{e^*(e)}$ is linear in e and conjugate linear in e^*.

At various points in the discussion of this paper, we shall be concerned with a real analytic manifold M_1 and real-analytic functions on M_1 with values in the operators from E to F, where F is another Banach space. There are several reasonable concepts of analyticity which may be used and which we discuss below. We shall establish that they are all equivalent, a fact which is well-known in the case of complex analytic (i. e. holomorphic) functions (cf. [10], pp. 92 ff.) where it is an easy consequence of the Cauchy integral formula. This result, however, does not carry over to the real analytic case without a different sort of argument, as follows from the fact that the corresponding conclusion is true in the complex case if E and F are any pair of locally convex separated linear topological spaces but does not hold without further structural assumptions in the real case (as can be exhibited by easy counter-examples).

DEFINITION. *Let M_1 be a real analytic manifold, E and F complex Banach spaces, T a function from M_1 into $L(E,F)$, the space of bounded* *linear operators from E to F. Let T_t be the operator assigned by T to the point t of M_1. Then:*

(i) *T is said to be weakly analytic on M_1 if for each e in E and each f^* in F^*, the function (T_te, f^*) is a real-analytic (complex-valued) numerical function on M_1.*

(ii) *T is said to be strongly analytic on M_1 if for each point t_0 of M_1 there exists a coordinate patch N on M_1 with coordinates $t = (t_1, \cdots, t_n)$ and t_0 having the coordinates $(0, \cdots, 0)$ and a constant $\epsilon > 0$ such that for all e in E*

$$(1) \qquad T_te = \sum_\alpha \frac{t^\alpha A_\alpha e}{\alpha!}; \qquad (t^\alpha = t_1^{\alpha_1} \cdots t_n^{\alpha_n}, \alpha = (\alpha_1, \cdots, \alpha_n)),$$

where A_α are elements of $L(E,F)$ independent of t and the series of (1) converges strongly in F for $|t| < \epsilon$.

(iii) *T is said to be uniformly analytic on M_1 if T is strongly analytic on M_1 and the series*

$$(2) \qquad T_t = \sum_\alpha \frac{t^\alpha}{\alpha!} A_\alpha$$

converges in the uniform topology on $L(E,F)$ for $|t| < \epsilon$, $\epsilon > 0$.

(iv) *T is said to be complex analytic on M_1 if the series (2) above converges for complex $(t_1, \cdots, t_n)$ with $|t| < \epsilon$.*

PROPOSITION 1. *For Banach spaces E and F the four definitions of analyticity above for functions from M_1 to $L(E,F)$ (i.e. weak analyticity, strong analyticity, uniform analyticity, and complex analyticity) are all mutually equivalent. In particular, every weakly analytic function on M_1 is a continuous function from M_1 to the norm topology on $L(E,F)$.*

Proof of Proposition 1. Obviously (iv) implies (iii), (iii) implies (ii), and (ii) implies (i). Hence we need only show that (i) implies (iv).

Suppose therefore that T is weakly analytic on M_1. For each e in E and f^* in F^*, the function $g_{e,f^*}(t) = (T_te, f^*)$ is analytic on M_1. Let t_0 be a point of M_1, N a coordinate patch on M_1 with coordinates $t = (t_1, \cdots, t_n)$ with $|t| < 1$ such that $t_0 = 0$. The function g_{e,f^*} has a power series expansion

$$(3) \qquad g_{e,f^*}(t) = \sum_\alpha \frac{t^\alpha}{\alpha!} a_\alpha(e, f^*)$$

which converges for $|t| < \epsilon_{e,f^*}$, where ϵ_{e,f^*} is a positive constant varying in general with e and f^*. The coefficient $a_\alpha(e,f^*)$ in the expansion (3) equals $(D^\alpha g_{e,f^*})(0)$, where $D^\alpha = \left(\frac{\partial}{\partial t_1}\right)^{\alpha_1}\cdots\left(\frac{\partial}{\partial t_n}\right)^{\alpha_n}$. It follows therefore that for each fixed α, $a_\alpha(e,f^*)$ is linear in e and conjugate linear in f^*. Since the expansion (3) converges for $|t| < \epsilon_{e,f^*}$, there exists a positive constant k_{e,f^*}, depending on e and f^*, such that for all α,

$$(4) \qquad |a_\alpha(e,f^*)| \leqq k_{e,f^*}{}^{|\alpha|}, \qquad (|\alpha| = \sum_{j=1}^{n} \alpha_j).$$

We may make k_{e,f^*} unique by choosing it as the infimum of all constants for which the inequality (4) holds.

Let B' be the closed subset of $E \times F^*$ which consists of the product $B \times B^*$ where B is the closed unit ball in E and B^* is the closed unit ball in F^*. We fix the coordinate patch N as above with the coordinate functions $(t_1,\cdots,t_n)$ and let $0 < \epsilon_0 < 1$. For each positive constant k, let

$$S_k = \{[e,f^*] : \| e \| \leqq 1, \| e^* \| \leqq 1, e \in E, f^* \in F^* \\ | D^\alpha(T_t e, f^*)(t) | \leqq k^{|\alpha|} \text{ for all } \alpha \text{ and } |t| < \epsilon_0\}.$$

For each $k > 0$, S_k is a closed subset of B' since if a sequence $[e_j, f^*_j]$ converges in B' to $[e_0, f^*_0]$, the sequence of functions $(T_t e_j, f^*_j)$ converges for each t in M_1 to $(T_t e_0, f^*_0)$ and the inequalities defining S_k pass over by an equicontinuity argument to the corresponding inequalities for the derivatives of $(T_t e_0, f^*_0)$. On the other hand, for each $[e,f^*]$ in B', the function $(T_t e, f^*)$ is analytic on the set $\{t : t \in R^n, |t| < 1\}$. Hence this function has a holomorphic extension to a complex neighborhood of the set $\{t : t \in R^n, |t| < 1\}$ in C^n. Since ϵ_0 is less than 1, there exists $d_0 > 0$ such that the power series expansion of $(T_t e, f^*)$ about each point t_1 with $|t_1| \leqq \epsilon_0$ has radius of convergence $\geqq d_0$ while on the complex d_0-neighborhood of the set $\{t : t \in R^n, |t| \leqq \epsilon_0\}$, the holomorphic extension of $(T_t e, f^*)$ is uniformly bounded by c_0. By the usual Cauchy integral estimates, it follows that $|D^\alpha(T_t e, f^*)(t)| \leqq c_0(d_0/2)^{-|\alpha|}$. Thus $[e, e^*]$ lies in at least one of the sets S_k, i. e. the sets S_k exhaust B'.

Let k range through the positive integers. The sequence of closed sets $\{S_k\}$ exhausts B'. By the Baire Category theorem, there exists at least one S_k containing an open subset of B'. Let S_k be such a set, and consider a closed ball B_0,

$$B_0 = \{[e,f^*] : \| e - e_0 \| \leqq \epsilon, \| f^* - f^*_0 \| \leqq \epsilon\},$$

which is entirely contained in S_k. Suppose that e_1 and f^*_1 are any pair of elements of E and F^*, respectively, with $\| e_1 \| \leqq \epsilon$, $\| f^*_1 \| \leqq \epsilon$. By the linearity of T_t, we have if we set $e = e_0 + e_1$, $f^* = f^*_0 + f^*_1$,

$$D^\alpha(T_te_1, f^*_1) = D^\alpha(T_t(e-e_0), (f^*-f^*_0))$$
$$= D^\alpha(T_te, f^*) - D^\alpha(T_te, f^*_0) - D^\alpha(T_te_0, f^*) + D^\alpha(T_te_0, f^*_0).$$

The for elements $[e,f^*]$, $[e,f^*_0]$, $[e_0,f^*]$, and $[e_0,f^*_0]$ all lie in the closed ball B_0 and hence in S_k. Hence,

$$|D^\alpha(T_te_1, f^*_1)(t)| \leqq 4k^{|\alpha|} \text{ for } |t| \leqq \epsilon_0, \|e_1\| \leqq \epsilon, \|f^*_1\| \leqq \epsilon.$$

It follows by homogeneity that for any pair e in E and e^* in E^*, we have

$$|D^\alpha(T_te, e^*)(t)| \leqq 4\epsilon^{-2}k^{|\alpha|}, |t| \leqq \epsilon_0.$$

Thus the sesquilinear form $D^\alpha(T_te, f^*)(t) = (A_{\alpha,t}e, f^*)$, where for each t and α, $A_{\alpha,t} \in L(E,F)$ and

$$\|A_{\alpha,t}\| \leqq 4\epsilon^{-2}k^{|\alpha|}, |t| \leqq \epsilon_0.$$

It follows immediately that the series

$$T'_t = \sum_\alpha \frac{t^\alpha}{\alpha!} A_{\alpha,0}$$

converges in the operator norm in $L(E,F)$ for all complex arguments $(t_1, \cdots, t_n)$ with $|t| < k^{-1}$. The sum of this series is an operator T'_t such that $(T'_te, f^*) = (T_te, f^*)$ for each e in B and f^* in B^* for $|t|$ sufficiently small. Since both (T'_te, f^*) and (T_te, f^*) are analytic for $|t| < k^{-1}$, they coincide for $|t| < k^{-1}$. Hence $T_t = T'_t$ for $|t| < k^{-1}$ and T_t is complex analytic on M_1. Q.E.D.

Remark 1. Since there is a natural identification of the Banach space E with the space $L(C^1, E)$ of bounded linear mappings of the complex number space C^1 into E, where C^1 is treated in the natural way as a one-dimensional complex Banach space, we may apply the results of Proposition 1 to the treatment of analytic functions with values in E itself. In all these cases, that of $L(E,F)$ and of E, we shall allow ourselves the privilege of speaking simply of analytic (i.e. real analytic) functions without qualification since the different senses of the term coincide by Proposition 1.

Remark 2. We remark explicitly that in the proof of Proposition 1, we have made use of the fact that a function T_t from M_1 to $L(E,F)$ which is weakly differentiable in the sense that (T_te, f^*) is differentiable in t for all e and f^* with the derivative bounded by $C\|e\|\|f^*\|$ must be strongly differentiable in $L(E,F)$. This follows for example from the results of [8].

Let H be a (complex) Hilbert space with inner product (u,v). We shall consider densely defined linear operators T in H, i.e. operators T with domain $D(T)$ dense in H and mapping that domain linearly on a range $R(T)$ in H. For each such operator, we designate by $N(T)$ the nullspace of T and by $G(T)$, the graph of T, the subspace of the product space $H \times H$ consisting of all pairs of the form $[u, Tu]$ for u in $D(T)$. T is said to be closed if $G(T)$ is closed in $H \times H$, closable if $T \subseteq T_1$ where T_1 is a closed operator in H. (By the relation $\subseteq$, we mean the usual "contained in" relation for densely defined operators, namely: T is contained in T_1 if $D(T) \subset D(T_1)$ while $Tu = T_1 u$ for $u \in D(T)$.) To each densely defined linear operator T, we assign a linear operator T^* called its adjoint with domain

$D(T^*) = \{v : (v, Tu)$ is a bounded linear functional of u in $D(T)\}$, and with values given by $T^*v = w$ if

$$(v, Tu) = (w, u), u \in D(T),$$

where the element w of H is uniquely determined by the latter equation because of the density of $D(T)$ in H. It is well-known that T^* is densely-defined if and only if T is closable, so that if T is closed, T^* is well-defined and indeed $(T^*)^* = T$. T^* is a closed operator, and moreover

$$N(T) = (R(T^*))^{\perp}, \quad N(T^*) = (R(T))^{\perp}$$

where $A^{\perp}$ denotes the orthogonal complement of A in H. Furthermore,

$$\mathrm{cl}(R(T)) = (N(T^*))^{\perp}, \quad \mathrm{cl}(R(T^*)) = (N(T))^{\perp}, \qquad (\mathrm{cl}(A) = \text{closure of } A \text{ in } H).$$

If $R(T)$ is closed, it follows that $R(T) = (N(T^*))^{\perp}$, and similarly if $R(T^*)$ is closed, then $R(T^*) = (N(T))^{\perp}$. Moreover, $R(T)$ is closed if and only $R(T^*)$ is closed. (For a detailed discussion of these and other properties of closed linear operators, we refer to the writer's paper [3].)

DEFINITION. *Let T_0 and T'_0 be two closed densely defined linear operators in H. Then T_0 and T'_0 are said to be formally adjoint to one another if $T' \subseteq (T_0)^*$, (or equivalently if $T_0 \subseteq (T'_0)^*$). If T_0 and T'_0 are a formally adjoint pair of operators, the closed operator T is said to be a realization of the pair (T_0, T'_0) if $T_0 \subseteq T \subseteq (T'_0)^*$.*

Obviously T is a realization of the pair (T_0, T'_0) if and only if T^* is a realization of the pair (T'_0, T_0).

DEFINITION. *T is said to be a solvable realization of the pair (T_0, T'_0) if T is a realization of the pair and in addition T^{-1} is defined and bounded on all of H.*

The definitions of realization and solvable realization are useful in formalizing the notion of boundary value problem and of solvable boundary value problem for a partial differential operator as was first pointed out by M. I. Visik in his Thesis [21]. Visik also remarked the following fact (cf. also Section 1 of [11]):

LEMMA. *For the formally adjoint pair of closed linear operators* (T_0, T'_0), *there exists a solvable realization* T *if and only if* $(T_0)^{-1}$ *and* $(T'_0)^{-1}$ *exist and are bounded on the ranges* $R(T_0)$ *and* $R(T'_0)$, *respectively.*

It is our purpose in the following discussion to establish sufficient conditions under which if $T_{0,t}$ and $T'_{0,t}$ are families of closed operators formally adjoint for each t in M_1, we can assert the existence of a solvable realization T_t of the pair $(T_{0,t}, T'_{0,t})$ such that $(T_{0,t})^{-1}$ is analytic in t on M_1.

Before proceeding to the statement and proof of our principal result, we shall first consider some preliminary results which will be useful in that proof.

LEMMA 1. *Let* H *be a Banach space,* W *a linear subset of* H *having a second Banach space structure on it such that the identity map of* W *into* H *is compact. Let* T *be a closed linear operator in* H *such that* $D(T) \subset W$. *Then* $R(T)$, *the range of* T, *is closed in* H.

Proof of Lemma 1. Let $G(T)$ be the graph of T. $G(T)$ is a closed subspace of $H \times H$, and hence itself a Banach space with respect to the norm $\| [u, Tu] \| = \| u \|_H + \| Tu \|_H$. Let J be the mapping of $G(T)$ into W given by $J([u, Tu]) = u$. J is well-defined on all of $G(T)$. and J is a closed linear operator, since if $[u_k, Tu_k] \to [u_0, Tu_0]$ and $u_k \xrightarrow[W]{} v$, then $u_k \to v$ in H, and $v = u_0$, i.e. $J([u_0, Tu_0]) = v$. Since $G(T)$ and W are both Banach spaces, it follows from the closed graph theorem that J is a continuous linear mapping, i.e. there exists a constant C such that

$$\| u \|_W \leqq C\{ \| u \|_H + \| Tu \|_H \}, u \in D(T). \tag{4}$$

Let $N(T)$ be the null-space of T, B the unit ball of $N(T)$ in the H norm. By the inequality (4), since $Tu = 0$ for all u in $N(T)$, it follows that $\| u \|_W \leqq C$ for all u in B. Since the identity map of W into H is compact, it follows that B is precompact in H and hence that $N(T)$ is of finite dimension.

Let P be a bounded idempotent linear operator mapping H onto $N(T)$. (Such operators exist by the Hahn-Banach Theorem.) Let $H_1 = (I - P)(H)$. H_1 is a closed subspace of H and if T_1 is the restriction of T to $D(T) \cap H_1$, then T_1 is densely defined in H_1, T_1 is a closed linear map from H_1 to H (since T_1 is closed in H), and $R(T_1)=R(T)$. To show that $R(T)$ is closed, it suffices to show $R(T_1)$ is closed, and for the latter it suffices to show that $(T_1)^{-1}$ is bounded (for then the extension of $(T_1)^{-1}$ by continuity must have a closed domain while such an extension must coincide with $(T_1)^{-1}$ itself since $(T_1)^{-1}$ along with T_1 is a closed linear map.)

To show that $(T_1)^{-1}$ is bounded, we remark first that $(T_1)^{-1}$ is well-defined since $N(T)\cap H_1=\{0\}$. Suppose that $(T_1)^{-1}$ is not bounded. Then there would exist a sequence $\{u_k\}$ in $D(T)\cap H_1$ such that

$$\| u_k \|_H = 1, \ \| Tu_k \|_H \to 0.$$

By the inequality (4), above, $\| u_k \|_W$ is uniformly bounded. By the compactness of the identity map of W into H, there must exist a subsequence of the sequence $\{u_k\}$ converging strongly in H. We may assume without loss of generality that this is the original sequence, so that $u_k \to u_0$ in H, $\| u_0 \|_H = 1$, $u_0 \in H_1$. On the other hand, $Tu_k \to 0$ in H. Since T is a closed operator, $u_0 \in D(T)$ and $Tu_0 = 0$. Thus $u_0 \in N(T)\cap H_1$ with $u_0 \neq 0$, yielding a contradiction. Hence $(T_1)^{-1}$ is bounded, and the proof of Lemma 1 is complete.

LEMMA 2. *Let M_1 be a real-analytic manifold, H a Hilbert space, R_0 a closed linear subspace of H, t_0 a point of M_1. Suppose that we are given a family $\{Q_t\}$ of bounded linear mappings of R_0 into H for t in M_1 such that Q_t is analytic in t on M_1 and $Q_{t_0}=0$. Let $P_t = J_0 + Q_t$, where J_0 is the identity map of R_0 into H. Then for t in some neighborhood N of t_0 on M_1, the range R_t of P_t is a closed subspace of H and the orthogonal projection operator E_t of H on R_t is analytic in t on N.*

Proof of Lemma 2. In the statement of Lemma 2, we have used the term "analytic" to denote any of the four equivalent types of analyticity previously discussed. By Proposition 1, we may assume that Q_t is continuous in the uniform topology at $t=t_0$ and it suffices to show that E_t is weakly analytic in t on a neighborhood N of t_0.

Since $Q_{t_0}=0$ and Q_t is continuous in norm, $\| Q_t \| \to 0$ as $t \to t_0$. Choose a neighborhood N_1 of t_0 such that for t in N_1, $\| Q_t \| < 1$. For t in N_1, P_t is one-to-one and $\| P_t u \| \geqq \| u \| - \| Q_t u \| \geqq (1 - \| Q_t \|) \| u \|$. Hence for such t, P_t is bicontinuous and R_t, its range, is closed in H.

Let E_t be the orthogonal projection of H on R_t in N_1, and set $E_0 = E_{t_0}$, the projection on R_0. Choose a smaller neighborhood N of t_0 such that $3 \| Q_t \| < 1$ for t in N. Let h be a fixed element of H. It suffices to show that $E_t h$ is analytic in t on N. For t in N, we know that $E_t h = P_t w_t$ for a uniquely determined w_t in R_0. Since P_t is analytic in t on N, it suffices to show that w_t is an analytic function in t on N. From the definition of orthogonal projection, we know that $(h - E_t h)$ is orthogonal to all of R_t. Hence for all w_1 in R_0, we have

$$(5) \qquad (h - E_t h, P_t w_1) = (h - P_t w_t, P_t w_1) = 0.$$

For each t, $P_t = J_0 + Q_t$, where J_0 and Q_t are bounded linear maps of R_0 into H. The adjoint mapping P_t^* mapping H into R_0 is given by $P_t^* = J_0^* + Q_t^*$, where $J^* = E_0$, the orthogonal projection on R_0, and $\| Q_t^* \| = \| Q_t \| < \frac{1}{3}$ for t in N. It follows from equation (5) that

$$(6) \qquad (P_t^* h - P_t^* P_t w_t, w_1) = 0 \text{ for all } w_1 \text{ in } R_0,$$

or, since the left-hand side of (6) lies in R_0, that

$$(7) \qquad P_t^* P_t w_t = P_t^* h$$

We obtain from equation (7), the following:

$$(8) \qquad (E_0 + Q_t^*)(J_0 + Q_t) w_t = P_t^* h,$$

or

$$(9) \qquad w_t + \{Q_t^* + Q_t + Q_t^* Q_t\} w_t = P_t^* h.$$

For t in N, however,

$$\| Q_t^* + Q_t + Q_t^* Q_t \| \leqq 3^{-1} + 3^{-1} + 9^{-1} < 1.$$

(We designate by Q_t^*, its restriction to the subspace R_0 of H as well as Q_t^* itself.) Hence w_t is uniquely determined by the Picard iteration method from equation (9). Moreover, since $(Q_t^* + Q_t + Q_t^* Q_t)$ is analytic in t, so is $(I + Q_t^* + Q_t + Q_t^* Q_t)^{-1}$ for t on N. Hence so is w_t, and $P_t w_t$, and the proof of Lemma 2 is complete.

We turn now to the principal result of the present section.

THEOREM 1.1. *Let H be a Hilbert space, M_1 a real analytic manifold. Suppose that we are given two families of densely defined closed linear operators in H, $\{T_{0,t}\}$ and $\{T'_{0,t}\}$, indexed on t in M_1. Let W and W' be two linear subsets of H each carrying a Banach space structure such that the identity maps of W and W' into H are compact. Suppose that:*

(1) *The pair $(T_{0,t}, T'_{0,t})$ are formally adjoint for each t in M_1.*
(2) *For all t in M_1, $D(T_{0,t}) = W$. For all t in M_1, $D(T'_{0,t}) = W'$.*
(3) *$T_{0,t}$ as a mapping of W into H depends analytically on t in M_1. $T'_{0,t}$ as a mapping of W' into H depends analytically on t in M_1.*
(4) *$N(T_{0,t}) = \{0\}$, $t \in M_1$. $N(T'_{0,t}) = \{0\}$, $t \in M_1$.*

Then there exists a family $\{T_t; t \in M_1\}$ *of closed linear operators in* H *such that*

(a) *For all* t *in* M_1, T_t *is a solvable realization of the pair* $(T_{0,t}, T'_{0,t})$, *i. e.* $T_{0,t} \subseteq T_t \subseteq (T'_{0,t})^*$ *and* $(T_t)^{-1}$ *is defined and bounded on all of* H.

(b) $(T_t)^{-1}$ *is analytic as a mapping from* M_1 *into* $L(H,H)$.

In addition, if for a fixed t_0 *in* M_1 *we are given a solvable realization* T_g *of the pair* (T_{0,t_0}, T'_{0,t_0}), *we can construct a family of solvable realizations* T_t *as above so that* $T_{t_0} = T_g$.

Proof of Theorem 1.1. We shall carry through the proof of Theorem 1.1 in several stages, beginning with the proofs of some crucial auxiliary results. The principal steps of the proof are the following:

(I) We show that $R(T_{0,t})$ and $R(T'_{0,t})$ are closed subspaces of H and that if E_t and E'_t are the orthogonal projections of H on $R(T_{0,t})$ and $R(T'_{0,t})$, respectively, then E_t and E'_t are analytic in t on M_1. It will follow that $(I - E_t)$ and $(I - E'_t)$ will give us analytic families of orthogonal projections on $N(T'_{1,t})$ and $N(T_{1,t})$, where we have set $T_{1,t} = (T'_{0,t})^*$, $T'_{1,t} = (T_{0,t})^*$.

(II) We construct a right inverse S_t for $T_{1,t}$ which is analytic in t on M_1. (By a right inverse, we mean of course that $T_{1,t}S_t = I$ for all t.)

(III) We modify S_t so that S_t agrees with $T_{0,t}^{-1}$ on $R(T_{0,t})$ for all t. Then T_t is defined to be S_t^{-1}.

(IV) We modify T_t so that $T_{t_0} = T_g$.

(I) (a) $R(T_{0,t})$ *and* $R(T'_{0,t})$ *are closed subspaces of* H. Since $D(T_{0,t}) = W$ and $D(T'_{0,t}) = W'$ where W and W' satisfy the hypotheses of Lemma 1, the conclusion follows from the conclusions of Lemma 1.

(I) (b) If E_t *is the orthogonal projection on* $R(T_{0,t})$, E'_t *the orthogonal projection on* $R(T_{0,t})$, *then* E_t *and* E'_t *are analytic in* t *on* M_1. We shall show that under the hypotheses of Theorem 1, we can derive the conclusion of (I) (b) from Lemma 2 above. Let t_0 be a fixed point in M_1. It suffices to show that E_t and E'_t are analytic in the neighborhood of t_0 in M_1. By the symmetry of hypotheses, it suffices to consider E_t alone. Let $R_0 = R(T_{0,t_0})$. R_0 is a closed subspace of H by the result of (I) (a). T_{0,t_0}^{-1} is well-defined on R_0 since $N(T_{0,t})$ is trivial for all t, and T_{0,t_0}^{-1} is a closed operator from R_0 into H since T_{0,t_0} is a closed operator in H. By the closed graph theorem,

T_{0,t_0}^{-1} is a continuous linear mapping of R_0 into H. However, it must also be a closed linear mapping into W since its range is contained in W and the identity map of W into H is compact and hence continuous. By the closed graph theorem once more, T_{0,t_0}^{-1} is a bounded linear map of R_0 into W. Define $P_t = T_{0,t}T_{0,t_0}^{-1}$. Since $D(T_{0,t}) = W$ for all t in M_1, P_t maps R_0 onto $R(T_{0,t})$ for all t, $P_{t_0} = I$, and since $T_{0,t}$ is an analytic family of linear mappings of W into H and T_{0,t_0}^{-1} is a continuous linear map of R_0 into W, P_t is an analytic family of bounded linear maps of R_0 into H. Setting $Q_t = P_t - I$, we see that the hypotheses of Lemma 2 are verified. Our conclusion follows from Lemma 2 since $R(P_t) = R(T_{0,t})$ for t in M_1.

(I)(c) *$(I - E_t)$ and $(I - E'_t)$ are analytic families of orthogonal projections on the spaces $N((T_{0,t})^*)$ and $N((T'_{0,t})^*)$, respectively i. e. on $N(T'_{1,t})$ and $N(T_{1,t})$ if we set $T'_{1,t} = (T_{0,t})^*$, $T_{1,t} = (T'_{0,t})^*$.* The analyticity of $(I - E_t)$ and $(I - E'_t)$ follows immediately from the analyticity of E_t and E'_t respectively. The range of $(I - E_t)$ is the orthogonal complement of the range of E_t, i.e. $(R(T_{0,t}))^{\perp}$, which equals $N((T_{0,t})^*)$, and similarly in the other case.

(II) *We may construct a right inverse S_t for $T_{1,t}$ which depends analytically on t in M_1.* We begin by constructing for each t in M_1 the bounded linear operator Y_t on H which is given by

$$Y_t = (T'_{0,t})^{-1}E'_t.$$

Since E'_t maps H onto $R(T'_{0,t})$ and since by the argument of (I)(b) above $(T'_{0,t})^{-1}$ is a bounded linear map of $R(T'_{0,t})$ into H, Y_t is a bounded linear operator on H for each t in M_1. Similarly Y_t is a bounded linear mapping of H onto W'. We wish to show that $t \to Y_t$ is an analytic function from M_1 to $L(H,H)$. Let h be any element of H. It suffices to show that Y_th is an analytic function from M_1 to H and, if t_0 is any given point of M_1, to show that Y_th is analytic in the neighborhood of t_0. Let $w'_t = Y_th$. Then $T'_{0,t}w'_t = E'_th$. For each t, w'_t lies in W' and hence may be written uniquely in the form $w'_t = (T'_{0,t_0})^{-1}(v_t)$ with v_t an element of $R'_0 = R(T'_{0,t_0})$. Since $(T'_{0,t_0})^{-1}$ is a continuous linear map, it suffices to show that v_t depends analytically on t in a neighborhood of t_0. But v_t satisfies the equation

$$T'_{0,t}(T'_{0,t_0})^{-1}v_t = E'_th$$

which may be written in the form

$$v_t + (T'_{0,t}(T'_{0,t_0})^{-1} - I)v_t = E'_th$$

Applying $E'_0 = E'_{t_0}$ to both sides of this last equation, we obtain

$$v_t + E'_0(T'_{0,t}(T'_{0t_0})^{-1} - I)v_t = E'_0E'_th.$$

The operators $E_0(T'_{0,t}(T_{0,t_0})^{-1} - I)$ and E_0E_t are analytic in t and the first has small norm for t near t_0. Hence for t near t_0, v_t can be obtained by the Picard iteration method from this equation and is analytic in t near t_0. Hence Y_t is analytic in t on M_1. It follows from Proposition 1 that Y_t^* is analytic in t on M_1. Set $S_t = Y_t^*$. We assert that S_t is a right inverse for $T_{1,t}$. Indeed, if $u = S_tv$,

$$(u, h) = (S_tv, h) = (v, Y_th)$$

Setting $h = T'_{0,t}w'$ for any w' in W', we find $Y_th = (T'_{0,t})^{-1}h = w'$, so that

$$(u, T'_{0,t}w') = (v, w')$$

Hence $u \in D(T_{1,t})$ and $T_{1,t}u = v$. Therefore S_t is a right inverse for $T_{1,t}$ depending analytically on t in M_1 and the proof of (II) is complete.

(III) *We may construct a right inverse S_t for $T_{1,t}$ which depends on t in M_1 and coincides with $(T_{0,t})^{-1}$ on $R(T_{,t})$.* We may *define $T_t = (S_t)^{-1}$, and T_t is the analytic family of solvable realizations sought in the first part of Theorem* 1.

By the result of (II), we have a right inverse S_t for $T_{1,t}$ already defined and depending analytically on t in M_1. We now modify it by defining

$$S^{\#}_t = (T_{0,t})^{-1}E_t + S_t(I - E_t).$$

The analyticity of the first summand in t is proved as in the proof of (II). The second summand is obviously analytic in t. For each t, $R(S_t)$ is obviously contained in $D(T_{1,t})$. We have moreover

$$T_{1,t}S^{\#}_t = T_{1,t}(T_{0,t})^{-1}E_t + T_{1,t}S_t(I - E_t) = E_t + (I - E_t) = I.$$

so that $S^{\#}_t$ is a right inverse for $T_{1,t}$. On $R(T_{0,t})$, $S^{\#}_t = (T_{0,t})^{-1}$. Since $S^{\#}_t$ is a right inverse for $T_{1,t}$, $(S^{\#}_t)^{-1}$ is defined and is a closed operator in H. Obviously $T_{0,t} \subseteq (S^{\#}_t)^{-1} \subseteq T_{1,t}$. Setting $T_t = (S^{\#}_t)^{-1}$, we have the family of solvable realizations T_t we seek.

(IV) *If T_g is a given solvable realization of the pair (T_{0,t_0}, T'_{0,t_0}) for a fixed t_0 in M_1, we can modify T_t so that it continues to satisfy the conditions of Theorem* 1 *while $T_{t_0} = T_g$. Let $S_0 = (T_g)^{-1}$.*

It suffices to construct S_t so that $S_{t_0} = S_0$. We remark that $S_{t_0} = (T_{0,t_0})^{-1} = S_0$ on $R(T_{0,t_0})$. Hence we need only modify S_{t_0} on the orthogonal complement of $R(T_{0,t_0})$, i. e. on $N(T'_{1,t})$.

S_{t_0} and S_0 both map $N(T'_{1,t_0})$, and since both are right inverses for T_{1,t_0}, it follows that for all h in $N(T'_{1,t_0})$, $Uh = (S_{t_0} - S_0)h = N(T_{1,t_0})$. Let $V = U(I - E_{t_0})$. V maps H continuously into $N(T_{1,t_0})$. We define $V_t = (I - E'_t)V$. For $t = t_0$, $V_{t_0} = V$. The family of operators V_t is analytic in t on M_1 and maps H for each t into $N(T_{1,t})$. We now define

$$S^{\#}_t = (T_{0,t})^{-1}E_t + (S_t - V_t)(I - E_t).$$

We verify as before that $S^{\#}_t$ is a right inverse for $T_{1,t}$, coincides with $(T_{0,t})^{-1}$ on $R(T_{0,t})$, and is analytic in t. In addition $S^{\#}_{t_0}$ coincides with S_0 on $N(T'_{1,t})$ and therefore coincides with S_0 on H. Hence if we set $T_t = (S^{\#}_t)^{-1}$, $T_{t_0} = (S_0)^{-1} = T_g$.

Thereby the proof of Theorem (1.1) is complete.

Remark. We may note without going into further detail that the conclusions and hypotheses of analyticity on Theorem 1.1 can be replaced by those of uniform continuity or differentiability and the argument given above for Theorem 1.1 remains valid.

In the later applications of our abstract results to obtain solutions of partial differential equations, it will be convenient to apply the following extension of Theorem 1.1 which is essentially a consequence of the latter.

THEOREM (1.2). *Let H be a Hilbert space, M_1 a real analytic manifold $\{T_{0,t}\}$ and $\{T'_{0,t}\}$ two families of closed linear operators in H indexed on t in M_1 which satisfy the hypotheses of Theorem 1.1. Let $T_{1,t} = (T'_{0,t})^*$. Suppose that we are given a family $\{v_t\}$ of elements of H depending analytically on t in M_1 and for a fixed t_0 in M_1, an element u_0 of $D(T_{1,t_0})$ such that $T_{1,t_0}u_0 = v_{t_0}$. Then there exists a family $\{u_t\}$ of elements of H depending analytically upon t in M_1 such that u_t lies in $D(T_{1,t})$ for $t \in M_1$, $T_{1,t}u_t = v_t$ for all t, while $u_{t_0} = u_0$.*

Proof of Theorem (1.2). By the second part of the conclusion of Theorem 1.1, it suffices to construct a solvable realization T_g of the formally adjoint pair (T_{0,t_0}, T'_{0,t_0}) such that $T_g^{-1}v_{t_0} = u_0$. Indeed, if this were done (so that we have an extension T of T_g over M_1 in the sense of Theorem 1.1 such that T_t for each t in M_1 is a solvable extension of $(T_{0,t}, T'_{0,t})$, $T_{t_0} = T_g$, and T_t^{-1} depends analytically on t), then the family $u_t = T_t^{-1}v_t$ would satisfy the conclusion of Theorem 1.2.

Let E_{t_0} be the orthogonal projection of H on the closed subspace $R(T_{0,t_0})$. We shall construct the right inverse S of T_g rather than T_g itself. S must coincide on $R(T_{0,t_0})$ with $(T_{0,t_0})^{-1}$. Let $v_1 = (I - E_{t_0})v_{t_0}$, $v_2 = E_{t_0}v_{t_0}$. Then $v_{t_0} = v_1 + v_2$, $Sv_{t_0} = Sv_1 + Sv_2$, and Sv_2 is a fixed element of $D(T_{0,t_0})$. Since

$Sv_{t_0} = u_0$, it follows that the only conclusion we impose upon S is that $Sv_1 = u_0 - (T_{0,t_0})^{-1}v_2 = w_0$ is a given element of $D(T_{1,t_0})$. Let S_1 be the inverse of any solvable realization of the pair (T_{0,t_0}, T'_{0,t_0}). We construct S from S_1 as follows:

$$Sv = (T_{0,t_0})^{-1}E_{t_0}v + S_1(I - E_{t_0})v + [w_0 - S_1v_1]\frac{(v, v_1)}{\|v_1\|^2}.$$

Since $(w_0 - S_1v_1)$ lies in $N(T_{1,t_0})$, it follows that S is the inverse of a solvable realization of the pair (T_{0,t_0}, T'_{0,t_0}). Furthermore, $Sv_1 = w_0$, as desired.

Q. E. D.

COROLLARY TO THEOREM (1.2). *Let H, M_1, $\{T_{0,t}\}$ and $\{T'_{0,t}\}$ be as in Theorem* 1.1. *Let $u_0 \in N(T_{1,t_0})$. Then there exists an analytic family u_t for t in M_1 with u_t lying in $N(T_{1,t})$ for each t in M_1.*

Proof of the Corollary. This is the special case of Theorem 1.2 with $v_t = 0$ for all t.

2. Differential operators in vector bundles. We turn to the consideration of linear systems of differential equations on a manifold using a general formulation for such equations along lines made familiar by Lions-Schwartz [16], Malgrange [17], and Kodaira-Spencer [13]. The advantage of this formulation for which the solutions sought are sections of a fibre bundle rather than systems of functions on the manifold is the obvious one that it contains the important case of equations for differential forms. Our notation differs somewhat from that, say, of Kodaira-Spencer, so that we make it explicit in the following preliminary discussion.

We consider a differentiable manifold M of dimension n which is connected, paracompact, and orientable (but not necessarily compact). Let B be a differentiable fibre bundle over M with projection map π whose fibre is a complex vector space C^p and whose structure group is the general linear group $GL_p(C)$. (By differentiable, we mean, throughout, differentiable of class C^∞.) Such a bundle B is said to be real-analytic if M is a real-analytic manifold and if the coordinate transformations of the bundle B are matrices b_{jk} with real-analytic coefficients. More explicitly, each point x_0 of M has a neighborhood N such that $\pi^{-1}(N)$ is homeomorphic to the product $C^p \times R^n$ by the mapping (ζ, x) where $x = (x_1, \cdots, x_n)$ is a system of coordinate functions on N and $\zeta = (\zeta_1, \cdots, \zeta_p)$ is a family of p complex-valued functions on $\pi^{-1}(N)$. If N' is another neighborhood on M together with another such product structure on $\pi^{-1}(N')$ given by (ζ', x'), then on $N \cap N'$, the coordinates x' are real-analytic functions of the coordinates x while for each x in $N \cap N'$, and the corresponding x', the points (ζ, x) and (ζ', x') coincide iff $\zeta'_j = \sum_k b_{jk}(x)\zeta_k$, with b_{jk} a given family of real-analytic functions on $N \cap N'$ depending on the pair of bundle coordinates.

A differentiable section of B over M is defined to be a differentiable mapping f of M into B such that $\pi(f(x)) = x$. More generally, we may speak of sections which are merely continouus or measurable, where by the latter we mean that on each of the coordinate neighborhoods N, f is a measurable mapping from R^n to $C^p \times R^n$ with respect to the usual Lebesgue measures, or more simply that each of the coordinates of these mappings is a measurable function with respect to the Lebesgue measure on R^n. If B is a real-analytic bundle, we may speak also of real-analytic sections since in that case both M and B are real-analytic manifolds.

Let $\mathcal{E}(B)$ be the linear space of all differentiable section of B over M. $\mathcal{E}(B)$ is a Fréchet space with respect to the natural topology of uniform convergence of the coordinates of sections and each of their derivatives over compact subsets of M. For each f in $\mathcal{E}(B)$, we define $\operatorname{Supp}(f)$ to be the smallest closed set F in M such that $f(x) = 0$ for x outside of F. Following L. Schwartz, $\mathfrak{D}(B)$ is defined to be the subfamily of $\mathcal{E}(B)$ consisting of sections f with compact support in M. For each compact subset K of M, $\mathfrak{D}_K(B)$ consists of f having their support in K, and each $\mathfrak{D}_k(B)$ is a closed subspace of $\mathcal{E}(B)$ and hence is a Fréchet space with respect to the induced topology. The space $\mathfrak{D}(B)$ is topologized as the inductive limit of the sequence $\mathfrak{D}_{K_s}(B)$, where $\{K_s\}$ is an increasing sequence of compact subsets exhausting M. Let $\mathfrak{D}'(B)$ and $\mathcal{E}'(B)$ be the strong dual spaces of $\mathfrak{D}(B)$ and $\mathcal{E}(B)$, respectively. In this context, the elements of $\mathfrak{D}'(B)$ correspond to the distributions of Schwartz in the usual scalar case, while those of $\mathcal{E}'(B)$ are the elements of $\mathfrak{D}'(B)$ having compact support, where $\operatorname{Supp}(\lambda)$ for $\lambda \in \mathfrak{D}'(B)$ is defined as the smallest closed set F in M such that $\lambda(f) = 0$ for each section f in $\mathfrak{D}(B)$ having its support contained in the complement of F. The spaces $\mathfrak{D}(B)$, $\mathcal{E}(B)$, $\mathfrak{D}'(B)$, and $\mathcal{E}'(B)$ are reflexive.

A linear differential operator with differentiable coefficients on B is defined to be a continuous linear mapping A of $\mathcal{E}(B)$ into $\mathcal{E}(B)$ such that $\operatorname{Supp}(Af) \subset \operatorname{Supp}(f)$, (i.e. a local operator). Let N be a coordinate patch corresponding to the bundle structure of B, (ζ, x) a set of bundle coordinates on $\pi^{-1}(N)$. Then each section f corresponds to a family of p functions $\zeta(x) = (f_1(x), \cdots, f_p(x))$ for x in R^n. In terms of these coordinates, the operator A may be written in the form (setting $(Af)_j$ as the coordinate of Af):

$$(Af)_j(x) = \sum_{k=1}^{p} \sum_{|\alpha| \leq r} a_{j,k,\alpha}(x) D^\alpha f_k(x).$$

Here we apply the standard notation for differential operators in R^n: $D_j = i^{-1}\dfrac{\partial}{\partial x_j}$, $\alpha = (\alpha_1, \cdots, \alpha_n)$ with the α_j's non-negative integers, $|\alpha| = \sum_{j=1}^{n} \alpha_j$, $D^\alpha = \prod_{j=1}^{n} D_j^{\alpha_j}$. The coefficients $a_{j,k,\alpha}(x)$, under our assumptions, are differentiable functions on N. We shall say that A has real-analytic coefficients if the bundle B is real-analytic and if the coefficients $a_{j,k,\alpha}(x)$ are real-analytic functions on N for each such coordinate neighborhood N.

Corresponding to a linear differential operator A in B and a given bundle coordinate system, we define the characteristic matrix of A by $A_{j,k}(x,\xi) = \sum_{|\alpha| \leqq r} a_{j,k,\alpha}(x)\xi^\alpha$ for each real n-vector ξ, with r the order of the differential operator in N (i. e. the largest integer such that $a_{j,k,\alpha}(x) \neq 0$ for some x in N and some α with $|\alpha| = r$.). *The differential operator A is said to be elliptic of order r if its order is r in every coordinate neighborhood N and if $A_{j,k}(x,\xi)$ is non-singular for every x in M and every $\xi \neq 0$ in R^n.* (The independence of the definition from the choice of the bundle coordinates on N is obvious.)

We suppose given a Hermitian inner product on $\mathcal{D}(B)$, written in the form (f,h), where in a given set of bundle coordinates over a neighborhood N in M, we have

$$(f,h) = \int_{R^n} \sum_{j,k} g_{j,k}(x) f_j(x) \bar{h}_k(x)\, dx,$$

for $\operatorname{Supp}(f) \subset N$, $\operatorname{Supp}(h) \subset N$, where f_j and h_j are the coordinates of f and h respectively, and $g_{jk}(x)$ is a differentiable function of x such that the matrix $\{g_{j,k}(x)\}$ is positive definite for each x. If B is analytic, we assume the $g_{jk}(x)$ analytic. Using this inner product, we may imbed $\mathcal{E}(B)$ in $\mathcal{D}'(B)$ by assigning to each h in $\mathcal{E}(B)$ the linear functional on $\mathcal{D}(B)$ given by

$$\lambda_h(f) = (f, f_0 h)$$

where f_0 is a differentiable function with compact support on M which equals 1 on a neighborhood of $\operatorname{Supp}(f)$. It follows by standard arguments that λ_h is independent of f_0 and that this assignment yields an injective continuous conjugate linear map of $\mathcal{E}(B)$ into $\mathcal{D}'(B)$. Its adjoint map yields a corresponding conjugate linear imbedding of $\mathcal{D}(B)$ into $\mathcal{E}'(B)$. We may thus speak as in the scalar case of elements of $\mathcal{D}'(B)$ lying in $\mathcal{E}(B)$, etc.

If B is a real-analytic bundle, let $\mathcal{A}(B)$ be the linear family of real-analytic sections of B over M. $\mathcal{A}(B)$ is given the structure of a locally convex topological vector space in the following way (see for example, Grothendieck [9], pp. 255, 315): For each compact subset K of M, we topologize the space $\mathcal{A}_K(B)$ of sections defined and analytic over a neighborhood of K and let $\mathcal{A}(B)$ be the intersection of the spaces $\mathcal{A}_K(B)$ with the weakest topology such that the injection mapping of $\mathcal{A}(B)$ into each $\mathcal{A}_K(B)$ is continuous. The topology on $\mathcal{A}_K(B)$ is given by considering $\mathcal{A}_K(B)$ as the inductive limit of the spaces $\mathcal{H}_R(K)$ where $\mathcal{H}_R(K)$ is defined with respect to a fixed finite covering of K by neighborhoods N_s carrying prescribed bundle coordinate systems by

$$H_R(K) = \{f\colon f \in \mathcal{A}_K(B),\ |D^\alpha f_j(x)| \leqq MR^{|\alpha|}\,|\alpha|!\ \text{on all}\ N_s \cap K\}.$$

For each $R > 0$, the space $H_R(K)$ is normed by taking as norm the least constant M above. (This amounts to taking for $H_R(K)$ all functions extendable to be holomorphic in a fixed neighborhood of K in a suitable complexification of B and taking as the topology on these functions the usual uniform norm topology on the extended holomorphic functions.) We shall have occasion to use below some of the properties of $\mathcal{A}(B)$ as a topological vector space which we shall consider in a later section.

Since A is a continuous linear maping of $\mathcal{E}(B)$ into itself and does not increase supports, A induces continuous linear mappings of each $\mathcal{D}_K(B)$ into itself for every compact subset K of M. Hence A induces a continuous linear map of $\mathcal{D}(B)$ into itself. The adjoint of this map which we denote by A' is a continuous linear map of $\mathcal{D}'(B)$ into itself, and we verify trivially that $\operatorname{Supp}(A'\lambda) \subset \operatorname{Supp}(\lambda)$ from the corresponding property of A. We wish to verify that A' is itself a linear differential operator (i.e. it carries $\mathcal{E}(B)$ continuously into $\mathcal{E}(B)$). This follows, however, from the fact that for f in $\mathcal{E}(B)$, h in $\mathcal{D}(B)$, we have $(A', f, h) = (f, Ah)$ by the definition of A'. Choosing f and h both to have their support in a neighborhood N on M carrying a famliy of bundle coordinates, we have

$$\begin{aligned}(A'f,h) &= \int_{R^n} \sum_{j,k} g_{jk}(x) f_j(x) \sum_{|\alpha|\leqq r} \bar{a}_{j,k,\alpha}(x)\,(\overline{D^\alpha h_k})(x)\,(dx\\ &= \int_{R^n} \sum_{j,k} \sum_{|\alpha|\leqq r} D^\alpha\{g_{j,k}(x)\bar{a}_{j,k,\alpha}(x) f_j(x)\}\bar{h}_k(x)\,dx\\ &= \int_{R^n} \sum_{k} \{\sum_{j} \sum_{|\beta|\leqq r} c_{j,k,\beta}(x)\,(D^\beta f_j)(x)\}\bar{h}_k(x)\,dx.\end{aligned}$$

Setting g^{jk} to be the inverse matrix of $g_{jk}(x)$, we find that the last integral may be written

$$\int_{R^n} \sum_{kl} g_{1k}(x)\{\sum_{j,m} g^{ml}(x)\,c_{j,m,\beta}(x)\,(D^\beta f_j)(x)\}\bar{h}_k(x)\,dx = (Cf, h),$$

where

$$(Cf)_j(x) = \sum_{|\beta| \leqq r} \sum_{l,m} g^{mj}(x)\, c_{l,m,\beta}(x)\, (D^\beta f_l)(x)$$
$$= \sum_{|\beta| \leqq r} \sum_{l} \{ \sum_{m} g^{mj}(x)\, c_{l,m,\beta}(x) \} (D^\beta f_l)(x)$$

defines a differential operator C. Thus $Cf \in \mathcal{E}(B)$, and the equation $(Cf,h) = (A'f,h)$ implies that $A'f = Cf$ on N. Thus A' is continuous from $\mathcal{E}(B)$ to $\mathcal{E}(B)$ and is itself a differential operator. As follows from the above calculation, in any given set of bundle coordinates $A'(x,\xi) = (\bar{A}(x,\xi))^{tr}$ so that A' is elliptic if and only if A is elliptic. If A has real-analytic coefficients, so does A'.

Since A', the formal adjoint of A, is itself a differential operator with differentiable coefficients as we have just observed, it follows that A' defines a continuous linear mapping of $\mathfrak{D}(B)$ into itself just as A did. We then define the action of our original differential operator A on $\mathfrak{D}'(B)$ as the adjoint of the mapping A' of $\mathfrak{D}(B)$ into itself. In other words, if $u \in \mathfrak{D}'(B)$, then Au is the unique element of $\mathfrak{D}'(B)$ which satisfies the relation

$$(Au, f) = (u, A'f)$$

for all f in $\mathfrak{D}(B)$.

Let $L^2(B)$ be the Hilbert space formed in the natural way from the measurable sections f of b for which $(f,f) = \| f \|^2 < +\infty$, or more precisely the equivalence classes of such measurable sections with respect to the equivalence relation of equality almost everywhere. (The last concept is taken locally with respect to Lebesgue n-measure and makes sense in the large because of the smoothness of the coordinate functions as well as the weights $g_{jk}(x)$.) The inner product in $L^2(B)$ is the inner product (f,h) defined above, which may be extended to measurable f and h in either of two ways: first by the local definition in terms of the $g_{jk}(x)$ which may be summed over a basis of the manifold after f and h have been written as sums of functions with small support, or second by considering $L^2(B)$ as the completion of the pre-Hilbert space $\mathfrak{D}(B)$ with respect to the same Hilbert space structure. In any case, it follows from the usual arguments that $\mathfrak{D}(B)$ is a dense subset of $L^2(B)$, or more precisely may be identified with a dense subset of $L^2(B)$. The identification mapping is continuous, injective, and has a dense range. Hence its adjoint induces a continuous, injective imbedding of $L^2(B)$ into $\mathfrak{D}'(B)$ such that the image is dense in $\mathfrak{D}'(B)$. Indeed as in the scalar case, $\mathfrak{D}(B)$ is itself dense in $\mathfrak{D}'(B)$.

If M' is a connected open subset of M, we shall denote by B' the restriction of the bundle B over M', i.e. the bundle induced over the manifold M' by the identity map of M' into M. Every element f of $\mathfrak{D}(B')$ may be identified with an element of $\mathfrak{D}(B)$ by setting $f=0$ outside of M'. In particular, every element f of $\mathfrak{D}(B')$ may be considered as an element of $L^2(B)$, and for any pair f and h of such elements (f,h) is then defined. Thus the inner product (f,h) in B induces a corresponding inner product on B' and moreover this extension procedure which we have just described yields an isometric map of a dense subset $\mathfrak{D}(B')$ of the Hilbert space $L^2(B')$ into $L^2(B)$. This mapping may be extended by continuity to an isometric imbedding of $L^2(B')$ into $L^2(B)$, an imbedding which is given by the same procedure of setting each section u in $L^2(B')$ equal to zero outside of M' to obtain a section in $L^2(B)$. The image of this imbedding consists of all sections in $L^2(B)$ which are equal to zero outside of M', or more strictly of all equivalence classes of such sections.

If M' is a precompact subset of M, we may define the Sobolev spaces on B' without ambiguity by setting for each $j \geqq 0$,

$W^{j,2}(B') = \{f : f \in L^2(B')$, In each bundle coordinate system (ζ, x) on a neighborhood N of a point of $\bar{M}'$, $D^\alpha f_j \in L^2{}_{loc}(N)$, for $|\alpha| \leqq j\}$.

In this definition, we let D^α denote the corresponding distribution differential operator in the Euclidean space on which N is mapped by the given coordinate system. If we choose a fixed finite family of bundle coordinate systems whose projections cover the closure of M' in M and let $f_{k,i}$ be the k-th coordinate of the section f in the i-th coordinate system over the i-th neighborhood N_i, we may introduce an inner product and Hilbert space structure in $W^{j,2}(B')$ by noting that on slightly shrunken neighborhoods $f_{k,i}(x)$ and $h_{k,i}(x)$ will have derivatives of order $\leqq j$ in L^2 rather than $L^2{}_{loc}$ and defining

$$(f,h)_{j,2,B'} = \sum_{j} \sum_{k} \sum_{|\alpha| \leqq j} \int_{R^n} D^\alpha f_{k,i}(x) \overline{D^\alpha h_{k,i}(x)}\, dx,$$

$$\| f \|^2{}_{j,2,B'} = (f,f)_{j,2,B'}.$$

The Hilbert spaces obtained for fixed M' and j from various choices of neighborhood families and bundle coordinate systems will all be mutually equivalent. The elements of $W^{0,2}(B')$ will coincide with those of $L^2(B')$. We denote by $W^{j,2}{}_{loc}(B)$ the family of sections f whose restriction to each precompact open subset M' of M lies in $W^{j,2}(B')$. For each $j \geqq 0$, $W^{j,2}{}_{loc}(B)$ is a Fréchet space with respect to its natural topology, namely the weakest topology sufficient to make all the restriction maps into $W^{j,2}(B')$ continuous for every precompact subset M'.

An important subspace of $W^{j,2}(B')$ is that which we designate by

$W^{j,2}{}_0(B')$, the closure in $W^{j,2}(B')$ of $\mathfrak{D}(B')$. For this subspace, we have as a very special case of the Sobolev Imbedding Theorem (cf. for example [2], Sec. 1) the following useful property:

LEMMA 2. *If M' is a precompact connected open subset of M, the identity map of $W^{j,2}{}_0(B')$ into $L^2(B')$ is a compact map.*

Let us now state the standard regularity theorems for linear elliptic systems of differential operators or, in our present terminology, for linear elliptic differential operators with differentiable coefficients on a vector bundle B.

LEMMA 2.2. *Let M be a connected differentable manifold which is paracompact and orientable, B a differentiable vector bundle over M, A a linear elliptic differential operator of order r on B, A' the formal adjoint of A. Then:*

(a) *If $u \in \mathfrak{D}'(B)$ and Au lies in $\mathcal{E}(B)$, then u itself lies in $\mathcal{E}(B)$.*

(b) *If $u \in \mathfrak{D}'(B)$ and Au lies in $W^{0,2}{}_{loc}(B)$, then u lies in $W^{r,2}{}_{loc}(B)$.*

(c) *If B is a real-analytic bundle and A has real-analytic coefficients and if for a given u in $\mathfrak{D}'(B)$, Au lies in $\mathcal{A}(B)$, then u itself lies in $\mathcal{A}(B)$.*

An important property of elliptic operators with real analytic coefficients, (as well as of some proper subclasses of elliptic operators with differentiable coefficients) is the property of uniqueness in the Cauchy problem in the small. This is the following:

$(U)_s$: *A is said to have the property of uniqueness in the Cauchy problem in the small with respect to B if for any connected open subset M' of M, B' the induced bundle, if u lies in $\mathcal{E}(B')$ with $Au = 0$ on M' while u vanishes on a non-trivial open subset of M', then u must vanish identically on M'.*

Indeed, for the case when A is elliptic and has real-analytic coefficients, such a section u must lie in $\mathcal{A}(B')$ by Lemma (2.2)(c), and, vanishing on a non-trivial open subset of M', must vanish identically.

DEFINITION 2.1. *Let A be a differential operator on B. Let A_{00} be the densely defined linear operator in $L^2(B)$ with domain $\mathfrak{D}(B)$ such that $A_{00}u = Au$ for $u \in D(A_{00})$. Similarly let A'_{00} be the corresponding operator for A'. Then:*

(a) *A_0, the minimal operator of A in $L^2(B)$, is defined to be the closure of A_{00} as an operator in $L^2(B)$.*

(b) *A'_0, the minimal operator of A' in $L^2(B)$, is defined to be the closure of A'_{00} as an operator in $L^2(B)$.*

(c) *A_1, the maximal operator of A in $L^2(B)$, is defined to be the Hilbert space adjoint of A'_{00} in $L^2(B)$, i.e. $A_1 = (A'_{00})^* = (A'_0)^*$.*

(d) *A'_1, the maximal operator of A' in $L^2(B)$, is defined by $A'_1 = A_{00}{}^* = A_0{}^*$.*

A closed linear operator T in $L^2(B)$ is said to be a realization of the differential operator A in $L^2(B)$ if $A_0 \subseteq T \subseteq A_1$, and a solvable realization if in addition T^{-1} is well-defined and defined on all of $L^2(B)$. (Similarly, the closed linear operator T' in $L^2(B)$ is said to be a realization of A' if $A'_0 \subseteq T' \subseteq A'_1$, and a solvable realization if in addition $(T')^{-1}$ is defined on all of $L^2(B)$.)

The existence of the closures considered in Definition (2.1) follows from the equality

$$(Af, h) = (f, A'h)$$

valid for all f and h in $\mathfrak{D}(B)$. Thus $A'_{00} \subseteq (A_{00})^*$, i.e. A_{00} has a densely defined adjoint and is therefore closable. The same argument applies to A'_{00}. In addition to A_0 and A'_0, their adjoints A'_1 and A_1 are closed operators since they are adjoint operators.

LEMMA 2.3. *Let M be a differentiable manifold, M' a precompact open subset of M. Suppose that B is a differentiable vector bundle over M, B' the restriction of B over M', A a linear elliptic differential operator of order r on B, A' its formal adjoint differential operator. Let A_0 and A'_0 be the minimal operators of A and A', respectively, in $L^2(B')$, A_1 and A'_1 the maximal operators of A and A', respectively, in $L^2(B')$. Then:*

(a) *The ranges of the operators A_0, A'_0, A_1, and A'_1 are closed in $L^2(B')$.*

(b) *The domains of A_0 and A'_0 coincide with $W^{r,2}{}_0(B')$.*

(c) *If A' satisfies the condition $(U)_s$ of uniqueness in the Cauchy problem in the small in B and $\mathrm{cl}(M') \neq M$, then $N(A'_0) = \{0\}$ and $R(A_1) = L^2(B')$. Similarly if A satisfies the condition $(U)_s$ on B, and $\mathrm{cl}(M') \neq M$, then $N(A_0) = \{0\}$ and $R(A'_1) = L^2(B')$. In particular if B is a real-analytic vector bundle over the real-analytic manifold M and if A has real-analytic coefficients, then $N(A_0) = N(A'_0) = \{0\}$ and $R(A_1) = R(A'_1) = L^2(B')$.*

Proof of Lemma 2.3. By Lemma (2.1), the identity mapping of

$W^{r,2}_0(B')$ into $L^2(B')$ is a compact map. By Lemma (1.1) of Section 1, if a closed linear mapping T in $L^2(B')$ has its domain contained in $W^{r,2}_0(B')$ then T has a closed range in $L^2(B)$. Thus if part (b) of the present Lemma were true it would follow that A_0 and A'_0 would have closed ranges in $L^2(B')$. But then so would their adjoints A'_1 and A_1. (This follows from the theorem that a closed operator T has a closed range if and only if T^* does, (cf. [3]). This last theorem is true in all Banach spaces and is trivial in Hilbert space.) Thus part (a) of our Lemma follows from part (b).

We now consider (b). If u is an element of $L^2(B')$ and lies in $D(A_0)$, there must exist a sequence $\{\phi_k\}$ from $\mathcal{D}(B')$ such that $\phi_k \to u$ in $L^2(B)$, $A\phi_k \to v$ in $L^2(B')$, where $v = A_0 u$. Extending u, v, and the ϕ_k to B by setting them equal to zero outside of M', we obtain the same convergence in $L^2(B)$. Thus every element u in $D(A_0)$ lies in the domain of the minimal operator of A in $L^2(B)$. Hence by part (b) of Lemma (2.2), u lies in $W^{r,2}_{loc}(B)$ and since M' is precompact, u must lie in $W^{r,2}(B')$. Consider the mapping J from $G(A_0)$ into $W^{r,2}(B')$. Since J is obviously closed, it is continuous by the closed graph theorem. Hence if $\phi_k \to u$, $A\phi_k \to A_0 u$, it follows that $\phi_k \to u$ in $W^{r,2}(B')$. Thus each u in $D(A_0)$ lies in the closure of $\mathcal{D}(B')$ in $W^{r,2}(B')$, i.e. $D(A_0) \subset W^{r,2}_0(B')$. On the other hand, if $\phi_k \to u$ in $W^{r,2}(B')$ it follows immediately from the fact that A is of order r that $A\phi_k \to v$ in $L^2(B')$ for some v. Hence $W^{r,2}_0(B') \subset D(A_0)$. Finally therefore $D(A_0) = W^{r,2}_0(B')$, and applying the same argument to A', part (b) of Lemma (2.3) is proved.

To prove part (c) of our Lemma, we first consider an element u of $N(A'_0)$. Extending u to be null outside of M', we have $u \in L^2(B)$, $A'u = 0$ on M, and $\operatorname{Supp}(u) \subset \operatorname{cl}(M')$ with $\operatorname{cl}(M')$ a proper subset of M. By Lemma (2.2) (a), u lies in $\mathcal{E}(B)$. By the condition $(U)_s$ on A', the solution u in $\mathcal{E}(B)$ of the homogeneous equation $A'u = 0$ which vanishes on $M - \operatorname{cl}(M')$ must be identically zero. Hence $N(A'_0) = \{0\}$. However, the orthogonal complement of $N(A'_0)$ in $L^2(B')$ is the closure of $R(A_1)$ which, by part (a) of our Lemma which we have already proved, equals $R(A_1)$ itself. Hence $R(A_1) = L^2(B')$. The conclusion if A satisfies $(U)_s$ follows by a symmetric argument. The final conclusion of (c) follows from Lemma (2.2) (c) which implies as we have already observed that for an elliptic operator A with real-analytic coefficients, both A and A' satisfy $(U)_s$.

We turn now to a simple proof of the Lax-Malgrange approximation theorem which it is one of the purposes of this paper to generalize.

PROPOSITION 2.1. *Let M be a differentiable manifold, M' an open subset of M such that $M - M'$ has no compact components. Suppose that B*

is a differentiable vector bundle over M, A a linear elliptic differential operator on B such that A' satisfies the condition $(U)_s$ on B. Let S be the family of solutions u of the equation $Au = 0$ in $\mathcal{E}(B)$, S_1 the family of solutions u_1 of the equation $Au_1 = 0$ in $\mathcal{E}(B')$. Then the restrictions of S to B' are dense in S_1 in the $\mathcal{E}(B')$-topology.

Proof of Proposition (2.1). We observe first that on S_1 the $\mathcal{E}(B')$ and the $L^2{}_{loc}(B')$ topologies are equivalent. Obviously the second is weaker than the first. Consider, on the other hand, the identity map of S_1 in its $L^2{}_{loc}(B')$ topology onto S_1 in its $\mathcal{E}(B')$ topology. The mapping is obviously closed. To show that it is continuous, it suffices by the closed graph theorem to know that S_1 is a Fréchet space in both topologies, or more precisely that it is closed in both $\mathcal{E}(B')$ and $L^2{}_{loc}(B')$. But a limit of u_k from S_1 in either topology would be an element v of $\mathcal{D}'(B')$ for which $A'v = 0$. Hence by Lemma (2.2)(a), v would lie in $\mathcal{E}(B')$ and therefore in S_1. Thus S_1 in either topology is a closed subspace of a Fréchet space, and the two topologies are equivalent on S_1.

It suffices by the preceding paragraph to prove that the restrictions of S to M' are dense in S_1 in the topology of $L^2{}_{loc}(B')$. It clearly suffices also to prove this only for M' which are also precompact. Consider an element w of $L^2{}_c(B')$, the subset of $L^2(B')$ of elements having compact support in M', such that $(w, u) = 0$ for all u in S. Suppose now that M itself is a precompact open subset of a manifold M'' with $\mathrm{cl}(M) \neq M''$ that B is the restriction of a bundle B'' on M'', and that A is defined on B''. Then extending w to an element of $L^2(B)$ by setting it equal to zero outside M', we obtain an element w of $L^2(B)$ orthogonal to S and in particular to $N(A_1)$ where A_1 is the maximal operator of A in $L^2(B)$. Under our assumption of the precompactness of M, it follows from Lemma (2.3) that the orthogonal complement of $N(A_1)$ is the closed subspace $R(A'_0)$ where A'_0 is the minimal operator of A' in $L^2(B)$. Hence there exists v in $D(A'_0)$ such that $A'_0 v = w$. Extending v to be zero in $M'' - M$, we may assume the same equation to hold in B''. Over $M'' - \mathrm{Supp}(w)$, v satisfies the equation $A'v = 0$ and therefore lies in $\mathcal{E}(M'' - \mathrm{Supp}(w))$. On $M'' - \mathrm{cl}(M)$ which is non-empty, v vanishes, and indeed the support of v is compact. It follows by an elementary topological argument that because of the assumption $(U)_s$ on A' in B'', the element v must have its support in B'. Hence for every element u_1 in S_1, $(w, u_1) = (A'v, u_1) = (v, Au_1) = 0$. However, $L^2{}_c(B')$ is the dual space of $L^2{}_{loc}(B')$, and we have shown under our special assumption on M that the annihilator of the restrictions of S annihilates the (possibly) larger subset S_1. Therefore the restrictions of S are dense in S_1.

To complete the argument, we must remove the assumption made upon M in the preceding paragraph. To do this, we construct an increasing sequence of open subsets $\{M_k\}$ of M such that $M_1 = M'$, each M_k has compact closure in its successor, $M_{k+1} - \mathrm{cl}(M_k)$ has no compact components, and the sequence exhausts M. Starting with u_1 which satisfies $Au_1 = 0$ over M_1, we may find a sequence $\{u_k\}$, by the preceding paragraph and by induction, with $Au_k = 0$ over M_k such that for $k \geqq 1$,

$$\| u_{k-1} - u_k \|_{L^2(B_{k-2})} < \epsilon/2^k$$

where $\epsilon > 0$ is given, B_k is the restriction of B over M_k, and M_0 is a given precompact subset of M'. Then $u = \sum_{k \geqq 1} (u_k - u_{k-1})\ (u_0 = 0)$, converges in $L^2(B_j)$ for every j, $Au = 0$ over each M_j, and

$$\| u - u_1 \|_{L^2(B_0)} < \sum_{k=1}^{\infty} 2^{-k}\epsilon = \epsilon.$$

The element u lies in S, and we have then shown that S is dense in S_1 in the topology of $L^2_{loc}(B')$. Q. E. D.

3. Differential operators depending upon a parameter. Compact manifolds. We now begin the study of families of differential operators depending differentiably or analytically upon a parameter by rederiving in a somewhat sharpened form the results of Kodaira and Spencer [13] on elliptic operators on a differentiable vector bundle over a compact manifold M. Our discussion is both simpler and more complete than that given in [13] and makes no assumption of strong ellipticity as is done in the latter. We consider also the sharper conclusions that can be drawn if the dependence of the elliptic operators in the parameter is real-analytic rather than merely differentiable.

Let M and M_1 be two connected paracompact differentiable manifolds with M orientable and compact. Let B be a differentiable vector bundle with projection π over $M \times M_1$, where following [13], we assume for simplicity that for a finite covering $\{N_k\}$ of M, each piece $\pi^{-1}(N_k \times M_1)$ has a product structure

$$\pi^{-1}(N_k \times M_1) = C^p \times N_k \times M_1.$$

A point on $\pi^{-1}(N_k \times M_1)$ is represented by the coordinates $(\zeta^{(k)}, x, t)$ with $t \in M_1$, and the point $(\zeta^{(k)}, x, t)$ coincides with $(\lambda^{(j)}, x, t)$ if and only if

$$\zeta_i^{(k)} = \sum_s b_{iskj}(x,t)\lambda_s^{(j)}$$

where the b_{iskj} are given differentiable functions of x and t. Let B_t be the restriction of the bundle B to the subspace $M \times \{t\}$ of $M \times M_1$, where $M \times \{t\}$ is identified in the natural way with M. We obtain then from the bundle B a family $\{B_t | t \in M_1\}$ of vector bundles over M. The bundle B or the associated family $\{B_t\}$ is called a differentiable family of vector bundles over M.

For each t in M_1, let A_t be a linear differential operator on B_t of a given order r independent of t. We assume that A_t depends differentiably upon t in the obvious sense, and further that we are given a Hermitian scalar product $(f,h)_t$ on $\mathcal{D}(B_t)$ which depends differentiably upon (x,t) in the sense that locally it is given by

$$\int_{N_k} \sum_{i,s} g_{isk}(x,t) f_i^{(k)}(x) h_s^{(k)}(x)\, dx$$

with the $g_{isk}(x,t)$ differentiable functions of (x,t).

We shall say that we are in the real-analytic case if both M and M_1 are real-analytic manifolds, B is a real-analytic vector bundle over $M \times M_1$, and the coefficients of the differential operator A_t and of the inner product $(f,h)_t$ are real-analytic functions of (x,t) on $M \times M_1$.

THEOREM 3.1. *Let M be a compact manifold, B_t a differentiable vector bundle over M for each t in M_1 with $\{B_t\}$ a differentiable family of vector bundles over M, A_t a linear elliptic differential operator of order r on B_t which is formally self-adjoint with respect to the inner product $(f,h)_t$. Then:*

(a) *If $A_{t,0}$ and $A_{t,1}$ are the minimal and maximal operators of A_t in $L^2(B_t)$, the two coincide, i.e. $A_{t,0} = A_{t,1}$. In particular the operator A_t restricted to $\mathcal{D}(B_t) = \mathcal{E}(B_t)$ is essentially self-adjoint in $L^2(B_t)$ i.e. its closure in $L^2(B_t)$ is a self-adjoint operator.*

(b) *If λ is a complex number with non-null imaginary part, A_t the unique closed realization of A_t in $L^2(B_t)$, then $(A_t - \lambda I)^{-1}$ exists and is a compact operator in $L^2(B_t)$. The operator A_t has a complete orthonormal set of eigenfunctions $\{e_{th}\}$ all of them lying in $\mathcal{E}(B_t)$, with real eigenvalues, $A_t e_{th} = \lambda_{th} e_{th}$, $\lambda_{th} \in R^1$. The real numbers λ_{th} for each fixed t have no finite accumulation point.*

Proof of Theorem 3.1. If M is compact, the spaces $W^{r,2}{}_{loc}(B)$ and $W^{r,2}{}_0(B)$ coincide. The latter is the domain of the minimal operator $A_{t,0}$. The former contains the domain of the maximal operator $A_{t,1}$. Hence $D(A_{t,1}) \subset D(A_{t,0})$ and it follows that $A_{t,1} = A_{t,0}$. Since A_t is formally self-adjoint in $L^2(B_t)$, $A'_{t,0} = A_{t,0}$, $A'_{t,1} = A_{t,1}$. Thus

$$A_{t,0} = A'_{t,0} = A'_{t,1} = (A_{t,0})^*,$$

from which it follows that $A_{t,0}$ is self-adjoint. Thus the proof of part (a) is complete.

For the proof of (b), we remark that since A_t is a self-adjoint operator, $(A_t-\lambda I)^{-1}$ will exist on all of $L^2(B_t)$ for $\mathrm{Im}(\lambda)=0$. The range of $(A_t-\lambda I)^{-1}$ is contained in $D(A_t)$ which equals $W^{r,2}{}_0(B_t)$. Consider the mapping J of u in $L^2(B_t)$ into $(A_t-\lambda I)^{-1}u$ in $W^{r,2}{}_0(B_t)$. This is obviously a closed linear map and hence by the closed graph theorem, J is a continuous map of $L^2(B_t)$ into $W^{r,2}{}_0(B_t)$. The map $(A_t-\lambda I)^{-1}$ of $L^2(B_t)$ into itself is the composition of J with the identity mop of $W^{r,2}{}_0(B_t)$ into $L^2(B_t)$. Since M is compact, the latter map is a compact map by Lemma (2.2). Hence $(A_t-\lambda I)^{-1}$ is compact. It follows immediately from the compactness of the resolvent $(A_t-\lambda I)^{-1}$ and the self-adjointness of A_t in $L^2(B_t)$ that the spectrum of A_t on $L^2(B_t)$ is a discrete subset of the real line R^1 that the corresponding eigenspaces are finite-dimensional, and span $L^2(B_t)$, and that eigenspaces corresponding to distinct eigenvalues are orthogonal. The eigenvectors e_{tk} are solutions of the homogeneous linear elliptic equation $(A_t-\lambda_k(t))e_{tk}=0$ and hence lie in $\mathcal{E}(B_t)$. Thus the proof of (b) and thereby of Theorem (3.1) is complete.

The remainder of the discussion of this Section like that of [13] is devoted to the study of the dependence of the $\lambda_k(t)$ as well as of the space S_t of solutions of $A_tu=0$ and a suitable generalized Green's operator G_t, all upon the variable t varying in M_1.

To state the results to be proved, some further definitions are necessary. If C is a rectifiable Jordan curve in the complex plane C^1 which contains 0 in its interest and such that no eigenvalues of A_t lie on C, we denote by $S_t(C)$ the closed subspace of $L^2(B_t)$ spanned by the eigenfunctions of A_t corresponding to eigenvalues lying in the interior of C and by $F_t(C)$ the orthogonal projection map of $L^2(B_t)$ on $S_t(C)$. It follows from the standard theory of self-adjoint operators that

$$F_t(C)=\frac{1}{2\pi i}\int_C(\lambda I-A_t)^{-1}d\lambda,$$

and that if we define the generalized Green's operator corresponding to C by

$$G_t(C)=\frac{1}{2\pi i}\int_C\lambda^{-1}(\lambda I-A_t)^{-1}d\lambda,$$

then

$$A_tG_t(C)=I-F_t(C).$$

Let S_t (without dependence on C) denote the closed subspace of solutions u

of $A_t u = 0$ in $L^2(B_t)$. Finally, if P_t for each t in M_1 is a bounded linear operator in $L^2(B_t)$, P_t is said to be differentiable in t iff $P_t\psi_t$ is differentiable for every differentiable section ψ over $M \times M_1$. Similarly in the analytic case, P_t is said to be real-analytic if $P_t\psi_t$ is real-analytic for every analytic section ψ. (In case the bundle B_t does not depend on t in M_1, this definition obviously coincides with the definitions of Section 1 by the theorem of that Section.)

THEOREM 3.2. *The eigenvalues $\lambda_k(t)$ depend continuously upon t, i.e. given an eigenvalue $\lambda_k(t_0)$ of A_{t_0} for t_0 in M_1 and $\epsilon > 0$, there is a neighborhood N of t_0 in M_1 such that for t in N, there exists an eigenvalue $\lambda_{k'}(t)$ which $|\lambda_k(t) - \lambda_{k'}(t)| < \epsilon$.*

THEOREM 3.3. *Let C be a rectifiable Jordan curve in C^1, M'_1 the open subset of t in M_1 such that A_t has no eigenvalues lying on C. Then $F_t(C)$ and $G_t(C)$ depend differentiably on t in M'_1.*

THEOREM 3.4. *The dimension of the linear space S_t is an upper semicontinuous function of t in M_1.*

THEOREM 3.5. *If* $\dim(S_t)$ *is independent of t in M_1, then for the operators F_t and G_t, where F_t is the orthogonal projection on S_t and $G_t = G_t(C)$ for a curve C enclosing at most the zero eigenvalue of A_t, F_t and G_t depend differentiably upon t in M_1.*

THEOREM 3.6. *If A_t is an analytic family of elliptic differential operators with analytic coefficients in an analytic family of bundles $\mathfrak{B}$, then the differentiable operator functions $F_t(C)$, $G_t(C)$, F_t, and G_t of Theorems (3.3) and (3.5) are analytic in the argument t upon the asserted domains of differentiability.*

Before proceeding to the proof of Theorems (3.2)-(3.6), we may follow the lead of [13], Lemma 1, by simplifying the formal situation slightly. Indeed, we may assume that for a small neighborhood of a point t_0 of M_1 (and all our assertions are upon the local character of functions, operator-valued and numerical), we may assume that the bundle $\mathfrak{B}_t$ is independent of t, or more precisely that the same family of bundle coordinate sets may be used on all B_t with coordinate transformations independent of t, that the elements of the Hilbert space $L^2(B_t)$ are a family of distributions independent of t and the inner product $(f, g)_t$ yields equivalent norms for t in the given neighborhood of t_0, and finally that $W^{r,2}(B_t)$ which is the domain of the differential operator A_t realized as an operator in $L^2(B_t)$ is independent of t.

Under this preliminary normalization, and letting $B = B_t$ for all t, let us now establish a key lemma from which the proofs of the above theorems will follow.

LEMMA 3.1. *If A_t is an injective mapping of $\mathcal{E}(B)$ into itself for t in M_1, then the linear operator A_t^{-1} is defined on all of $\mathcal{E}(B)$ and is differentiable in t. If A_t and B are analytic, so is A_t^{-1}.*

One should compare Lemma (3.1) with Proposition 1 of [13], pp. 51-53, where the same conclusion is obtained under the hypothesis that A_t is surjective for all t in M_1 and that there exists a constant $c > 0$ independent of t such that $\| A_t f \|_{L^2(B)} \geqq c \| f \|_{L^2(B)}$.

Before proceeding to the proof of Lemma (3.1), we introduce the following sharper version of Lemma (2.2) which like the latter follows from the standard interior regularity theory for elliptic systems of differential equations. (For a unified and complete presentation of this theory for the case of a single scalar equation, we refer to the writer's recent paper [3]. The arguments given there go over with only formal changes to the case of systems.)

LEMMA 3.2. *Let B be a p-vector bundle over the compact manifold M, A a linear elliptic differential operator of order r on B. For each $j \geqq 0$, if $u \in \mathcal{D}'(B)$ and $Au \in W^{j,2}(B)$, then u lies in $W^{j+r,2}(B)$.*

Proof of Lemma (3.1). We shall prove Lemma (3.1) in two parts:

(a) the differentiable case,

(b) the analytic case.

Let us consider the differentiable case first. We shall denote by A_t not only the differential operator A_t acting either on $\mathcal{E}(B)$ or $\mathcal{D}'(B)$ but also the unique realization of A_t in $L^2(B)$. The operator A_t referred to in the statement of Lemma (3.1) is, of course, the operator on $\mathcal{E}(B)$ but we note first that by Lemma (2.2), that if for u in $\mathcal{D}'(B)$, we have $A_t u$ in $\mathcal{E}(B)$, then u itself lies in $\mathcal{E}(B)$.

We assert first that for each t in M_1, the operator A_t on $L^2(B)$ has a bounded inverse A_t^{-1} defined on all of $L^2(B)$. By the earlier hypotheses of of this section, for each t in M_1, there exists an inner product $(f, g)_t$ on $L^2(B)$ with respect to which A_t is formally self-adjoint. By Theorem (3.1), we know that A_t is actually self-adjoint on the Hilbert space $L^2(B)$ with the t-inner product, that A_t has its spectrum on the real axis and that its spectrum is discrete and eigenspaces finite-dimensional, and that the resolvent $(A_t - \lambda I)^{-1}$ is a compact mapping in $L^2(B)$ for $\operatorname{Im}(\lambda) \neq 0$. It follows from the theory of operators with compact resolvents that the range of A_t is closed and is the orthogonal complement of the null-space of A_t in $L^2(B)$. However, if u in $L^2(B)$ is a solution of $A_t u = 0$, then $u \in \mathcal{E}(B)$ and by hypothesis, $u = 0$. Thus $R(A_t) = L^2(B)$ and A_t^{-1} is well-defined and hence bounded on all of $L^2(B)$ for each t in M_1.

Let $f \in W^{j,2}(B)$, $u_t = A_t^{-1} f$. By Lemma (3.2), each u_t lies in $W^{j+r,2}(B)$ for t in M_1. We propose to show that the maps $t \rightarrow u_t$ is a differentiable function from M_1 to $W^{r+j,2}(B)$. It will then follow that for f in (B), $A_t^{-1} f$ is a differentiable function from M_1 to $\mathcal{E}(B)$ since by the Sobolev Imbedding Theorem the $W^{j,2}$-norms form a family of norms defining the topology of $\mathcal{E}(B)$. (See [2] for a proof of the Sobolev Theorem in the scalar case, of which the p-vector case is a trivial variant.) We wish to show that $u_t = A_t^{-1} f$ is differentiable in $W^{r+j,2}(B)$ for f in $W^{j,2}(B)$. However, the correspondence $t \rightarrow A_t$ is a differentiable map of M_1 into the space of continuous linear mappings of $W^{r+j,2}(B)$ into $W^{j,2}(B)$. Each of these mappings A_t between this pair of domain and range spaces is both injective and surjective, and hence bi-continuous. It is a well-known fact of Banach space theory that if L_t is a differentiable family of bicontinuous linear maps of a Banach space X on a Banach space Y, then the family L_t^{-1} is also differentiable. Applying this fact to our case, we obtain the differentiability of u_t as a function into $W^{r+j,2}(B)$. Thereby, the proof of the differentiable case is complete.

We turn now to the analytic case where M, M_1, and B are all analytic and the coefficients of A_t depend analytically jointly upon all their variables. Let f_t be an analytic function from M_1 to $\mathcal{A}(B)$, $u_t = A_t^{-1} f_t$. We must show that u_t yields an analytic function from M_1 to $\mathcal{A}(B)$. Without loss of generality we may assume that M_1 is a complex analytic manifold and all the data are complex analytic instead of real-analytic, and indeed we may assume for the sake of simplicity that M_1 is a polycylinder in C^s. It suffices to consider a coordinate neighborhood N on M and a bundle coordinate neighborhood in B above N and to show that the components of u_t in the latter are analytic functions of (x, t) for x in N, t in M_1. We may therefore assume that A_t is an elliptic system of differential operators of order r and rank p acting on p-vector functions on M. By the holomorphic analogue of the argument of the preceding paragraph, $t \rightarrow u_t$ is holomorphic from M_1 to $W^{r+j,2}(B)$. Considered as a p-vector function on $M \times M_1$, therefore u_t satisfies the new elliptic system

$$(A'_t A_t + \Delta^r I_p) u_t = A'_t f_t,$$

where I_p is the identity matrix of rank p, Δ is the Laplace operator on M_1;

and A'_t is the formal adjoint of A_t with respect to Lebesgue n-measure on N. Since $A'_t f_t$ is analytic in (x, t) while the new elliptic system of order $2r$ and rank p has coefficients which are analytic on $N \times M_1$, it follows that u_t is analytic in (x, t) and the treatment of the analytic case is complete, and thereby the proof of Lemma (3.1).

We complete this section by giving the brief and simple proofs of Theorems (3.2) to (3.6).

Proof of Theorem (3.2). Let λ_0 be an eigenvalue of the differential operator A_0 corresponding to a parameter value t_0 in M_1. Denote by $\| u \|_{r,2}$ the norm in $W^{r,2}(B)$, and by $\| v \|$ without subscripts the norm in $L^2(B)$. From Lemma (2.2) and the closed graph theorem, it follows that there exists a constant C such that for all u in $W^{r,2}(B)$

$$\| u \|_{r,2} \leqq C\{\| A_0 u \| + \| u \|\}.$$

Choose a rectifiable Jordan curve C around λ_0 containing no other eigenvalues of A_0 in its closure, and for v in $L^2(B)$, set $u_\lambda = (\lambda I - A_0)^{-1} v$. Then by the above inequality

$$\| u_\lambda \|_{r,2} \leqq C\{\| A_0 u_\lambda \| + \| u_\lambda \|\} \leqq C\{\| v \| + (1 + |\lambda|) \| u_\lambda \|\}.$$

Since $(\lambda I - A_0)^{-1}$ is uniformly bounded for λ on C as is $|\lambda|$, it follows that $(\lambda I - A_0)^{-1}$ is a uniformly bounded family of linear maps of $L^2(B)$ into $W^{r,2}(B)$ for λ on C. Since A_t is a continuous family of linear maps of $W^{r,2}(B)$ into $L^2(B)$, it follows that for t near t_0, $(\lambda I - A_t)^{-1}$ exists for λ on C and is continuous in λ and t. Hence

$$F_t(C) = (2\pi i)^{-1} \int_C (\lambda I - A_t)^{-1} d\lambda$$

is a continuous family of projections for t near t_0 whose range consists of the eigenspaces of A_t within C. Since $F_t(C)$ has non-null dimension for $t = t_0$, it must have non-null dimension for t near t_0 since it is continuous in t. Hence A_t has an eigenvalue within C and Theorem (3.2) is proved.

Proof of Theorem (3.3). Let C be a rectifiable Jordan curve in the complex plane, t_0 a parameter value such that the corresponding differential operator A_0 has no eigenvalues on C. By the argument of the proof of Theorem (3.2), for t near t_0, $(\lambda I - A_t)^{-1}$ exists and is uniformly bounded. Consider

$$F_t(C) = (2\pi i)^{-1} \int_C (\lambda I - A_t)^{-1} d\lambda,$$

$$G_t(C) = (2\pi i)^{-1} \int_C \lambda^{-1} (\lambda I - A_t)^{-1} d\lambda.$$

Without changing the value of $G_t(C)$ or $F_t(C)$ for t near t_0, we may replace C by a differentiable or even by a real-analytic curve. Then the differential operator $(A_t - \lambda I)$ has no null space for λ on C, t near t_0, and has coefficients depending differentiably on (t, λ). Hence $(\lambda I - A_t)^{-1}$ and $\lambda^{-1}(\lambda I - A_t)^{-1}$ depend differentiably on (t, λ) and, thereby, $F_t(C)$ and $G_t(C)$ depend differentiably on t.

Proof of Theorem (3.4). Let A_0 be the differential operator corresponding to t_0 and let C be a restifiable Jordan curve about zero in the complex plane containing no other eigenvalues of A_0. For t near t_0, $F_t(C)$ is defined and continuous. For each t, $F_t(C)$ is a projection on the eigenspaces of A_t corresponding to eigenvalues within C. The dimension of the range of $F_t(C)$ is constant for t near t_0 by the continuity of $F_t(C)$ in t, its dimension for $t = t_0$ is $\dim(S_0)$ where S_0 is the null-space of A_0, and its dimension for all t is $\geqq \dim(S_t)$. Hence $\dim(S_0) \geqq \dim(S_t)$. Q. E. D.

Proof of Theorem (3.5). It follows from the proof of Theorem (3.4) that if $\dim(S_t)$ is constant for t near t_0, then for the curve C used in the latter proof for t near t_0 the only eigenvalue of A_t within C is at zero. Hence $F_t = F_t(C)$, $G_t = G_t(C)$, and the assertion of Theorem (3.5) follows from that of Theorem (3.3).

Proof of Theorem (3.6). We need only repeat the proofs of Theorems (3.3) and (3.5) using the result of Lemma (3.1) for the analytic case.

4. Analytic families of operators and kernels. We now turn back to the general framework of Section 2 with B a real-analytic vector bundle over the analytic manifold M and under the assumption that M is *not necessarily compact.* (This last assumption is made in distinction from that of Section 3 where M was assumed compact and the assumption in Sections to follow that M is definitely non-compact.)

Let M_1 be a second real-analytic manifold serving as the parameter space for an analytic family $\{A_t\}$ of elliptic differential operators of order r in B. Suppose that for each t in M_1, we have a solvable realization T_t in $L^2(B)$ with $R_t = T_t^{-1}$ depending analytically upon t in M_1. As we shall describe in more detail below, each bounded operator R in $L^2(B)$ has a kernel representation by the Schwartz kernel theorem ([19]). It is our object in the present Section to discuss the analyticity properties of the kernels associated with the operators R_t and, more generally of the sections $T_t^{-1} f_t$ where f_t is an analytic mapping from M_1 into $\mathfrak{A}(B)$. In Sections 5 and 6, we shall apply the results so obtained to the proofs of the existence of fundamental solutions depending analytically on t and of the analytic solvability of elliptic equations with coefficients and inhomogeneous terms varying analytically with t.

By a simple variant of the usual Schwartz kernel theorem, there corresponds to each continuous mapping of $\mathfrak{D}(B)$ into $\mathfrak{D}'(B)$ an unique kernel in the following sense: Let $B \otimes B$ be the analytic vector bundle over $M \times M$ with fibre the tensor product of the fibre of B with itself, with coordinate neighborhoods in $B \otimes B$ given for any pair of coordinate neighborhoods in B, (ζ, x) and (ζ', x'), by $(\zeta \times \zeta'. (x, x'))$, and with coordinate transformations given as the tensor products of the corresponding coordinate transformations in each factor. We may form the spaces $\mathfrak{D}(B \otimes B)$ and $\mathfrak{D}'(B \otimes B)$, and identify the latter with the space of continuous linear mappings of $\mathfrak{D}(B)$ into $\mathfrak{D}'(B)$ by assigning to the element K of $\mathfrak{D}'(B \otimes B)$, the mapping of $\mathfrak{D}(B)$ into $\mathfrak{D}'(B)$ which maps f in $\mathfrak{D}(B)$ into $K(f)$ given by

$$\langle K(f), g\rangle = \langle K, f \times g\rangle, g \in \mathfrak{D}(B).$$

This identification is a bicontinuous linear mapping between the strong topology on $\mathfrak{D}'(B \otimes B)$ and the topology of uniform convergence on compact subsets in $\mathcal{L}(\mathfrak{D}(B), \mathfrak{D}'(B))$. In particular, every analytic function from M_1 to $\mathcal{L}(\mathfrak{D}(B), \mathfrak{D}'(B))$ corresponds to an analytic function from M_1 to $\mathfrak{D}'(B \otimes B)$.

If R is a bounded linear mapping of $L^2(B)$ into itself, R obviously determines a continuous linear mapping of $\mathfrak{D}(B)$ into $\mathfrak{D}'(B)$ by composing R with the natural injection of $\mathfrak{D}(B)$ into $L^2(B)$ on one side and the natural imbedding of $L^2(B)$ into $\mathfrak{D}'(B)$ on the other. Hence R determines a kernel K, and since the induced map of $\mathfrak{D}(B)$ into $\mathfrak{D}'(B)$ determines R, K is uniquely determined by R and uniquely determines R. The assignment of kernels to mappings in this fashion is a continous map of $\mathcal{L}(L^2(B), L^2(B))$ into $\mathfrak{D}'(B \otimes B)$. In particular if the operators R_t are the inverses of a family of realizations of the corresponding differential operators A_t and if the operators R_t vary analytically with t on M_1 as operators in $L^2(B)$, then their kernels $\{K_t\}$ yield an analytic function from M_1 to $\mathfrak{D}'(B \otimes B)$.

THEOREM 4.1. *Let B be an analytic vector bundle over the real-analytic manifold M, M_1 a second real-analytic manifold, $\{A_t\}$ a family of elliptic differential operators of a fixed of order r in B with analytic coefficients with A_t analytic in t on M_1. Let $\{T_t\}$ be a family of solvable realizations of A_t in $L^2(B)$, $t \in M_1$, such that if $R_t = T_t^{-1}$, R_t is an analytic family of operators in $L^2(B)$.*

Then: If K_t *is the kernel of* R_t, *each* K_t *lies in* $L^q{}_{loc}(B\otimes B)$ *for* $q^{-1}>1-n^{-1}r$, *and the mapping* $t\to K_t$ *is an analytic mapping of* M_1 *into* $L^q{}_{loc}(B\otimes B)$. *Furthermore* K *is analytic over* $(M\times M-\Delta_0)\times M_1$, *where* Δ_0 *is the diagonal of* $M\times M$.

THEOREM 4.2. *Let* B *be an analytic vector bundle over the analytic manifold* M, $\{A_t\}$ *a family of elliptic differential operators of order* r *in* B *with analytic coefficients varying analytically with* t *in another analytic manifold* M_1. *Suppose that* T_t *is a solvable realization of* A_t *in* $L^2(B)$ *for each* t *such that* T_t^{-1} *depends analytically upon* t *in* M_1. *Let* f *be an analytic function from* M_1 *to* $\mathcal{a}(B)$ *and set* $u_t=T_t^{-1}f_t$. *Then* u *is an analytic function from* M_1 *to* $\mathcal{a}(B)$, *and indeed* u *is analytic over* $M\times M_1$.

To prove Theorem (4.1), we use the following variant of the Dunford-Pettis theorem ([7]) rather than the Schwartz Kernel theorem.

LEMMA 4.1. *Every continuous linear mapping* R *of* $L^p(B)$ *into* $L^\infty(B)$, $(1\leqq p<+\infty)$, *may be written as a integral transform of the form*

$$(Rf)(x)=(f(\cdot),K_x,\cdot),\ \text{a.e. in } x$$

where K *is an uniquely determined element of* $L^{p'}{}_{loc}(B\otimes B)$ *such that*

$$\| R\|^{p'}=\text{ess.sup.}_x\| K_x,\cdot\|^{p'}_{L^{p'}(B)}<+\infty.$$

The proof of Lemma (4.1) from the standard form of the Dunford-Pettis theorem for scalar functions is routine.

To apply Lemma (4.1) in the proof of Theorem (4.1), we shall use the strengthened form of our previously applied regularity Lemma (2.1) of Section 2 which gives the corresponding result for general p with $1<p<+\infty$ rather than just $p=2$.

LEMMA 4.2. *Let* A *be an elliptic differential operator of order* r *with differentiable coefficients on the differentiable vector bundle* B *over* M. *Let* u *be an element of* $\mathfrak{D}'(B)$ *with* Au *in* $L^p(B)$ *for some* p *with* $1<p<\infty$. *Then* u *lies in* $W^{r,p}{}_{loc}(B)$.

For the case of scalar functions a proof of Lemma (4.2) is given in Section 2 of [3]. The corresponding result in our case, which is essentially the case of an elliptic system on an open subset of E^n, may be obtained by the standard trick of replacing the system $\{A_j\}$ by the "almost-diagonalized" system

$$\sum_j Q_jA_j=\text{Det}(A)\cdot I^p+S$$

where Det (A) is the suitably defined determinant of the system, I^p is the unit p-matrix, and S is a system of lower order than Det(A). The conclusion of Lemma (4.2) for the elliptic system $\{A_j\}$ then follows from the standard interior L^p-estimates for the scalar elliptic differential operator Det(A).

The combination of Lemmas (4.1) and (4.2) is made possible by one part of the Sobolev Imbedding theorem (cf. Section 1 of [2], Lemma 5), which we state formally for the sake of reference:

LEMMA (4.3). *If $p^{-1} < n^{-1}r$, then every element of $W^{r,p}{}_{loc}(B)$ lies in $C^0(B)$, and the imbedding mapping is continuous.*

Combining Lemmas (4.1), (4.2), and (4.3), we obtain the following

PROPOSITION (4.1). *Let A be an elliptic differential operator of order r with differential coefficients on the differentiable vector bundle B over the manifold M. Let T be a solvable realization of $A, R = T^{-1}$. Then there exists a kernel $K \in L^{p'}{}_{loc}(B \otimes B)$, for any p' such that $(1/p') > \frac{n-r}{n}$, such that*

$$(Rf)(x) = (f, K_x(\cdot)),$$

for almost all x in M. If a family of operators $\{T_t\}$ have $\| T_t^{-1} \|$ bounded, the corresponding family of kernels is bounded in $L^{p'}{}_{loc}(B \otimes B)$.

Proof of Theorem (4.1). By Proposition (4.1), K_t lies in $L^q{}_{loc}(B \otimes B)$ for each t in M_1 $(q = p')$. Without loss of generality, we may assume that M_1 is a complex analytic manifold (since our conclusions are local in character) that $t \to R_t$ is a holomorphic function from M_1 into $\mathcal{L}(L^2(B), L^2(B))$ and that the coefficients of A_t are holomorphic in t on M_1. We remark in addition that in this extension to the complex which we have just implicitly performed, it remains true that for each f in $L^2(B)$,

$$A_t R_t f = f.$$

By the Schwartz Kernel theorem, K_t gives an holomorphic function from M_1 to $\mathcal{D}'(B \otimes B)$. If we restrict ourselves to a suitable neighborhood of the original real manifold M_1 in the new complex analytic manifold by which we have replaced it, all the operators A_t are still of order r and are all elliptic. Hence all the K_t lie in $L^{p'}{}_{loc}(B \otimes B)$ and the set $\{K_t, t \in C\}$ is bounded in $L^{p'}{}_{loc}(B \otimes B)$ for any compact subset C of M_1. The conclusion of Theorem (4.1) then follows from Lemma (4.4) below, if the latter is applied to the pair of spaces $E = \mathcal{D}'(B \times B)$ and $E_1 = L^{p'}(B \times B)$.

LEMMA (4.4). *Let h be a holomorphic function from M_1 to a locally convex space E with the property that the values taken on by h lie in a linear subset E_1 of E carrying a second stronger locally convex topology. Suppose that E_1 is complete in its topology and that for each compact subset C of M_1, the set $\{h(t), t \in C\}$ is bounded in E_1. Then h is holomorphic from M_1 to E_1.*

Proof of Lemma (4.4). Since the hypothesis and conclusions are local in character, we may asume that M_1 is a subset of complex n-space, that C is a polycylinder with distinguished boundary B whose closure is contained in M_1, and prove that the function h is holomorphic from the interior of C into E_1. For points t inside C, we have the Cauchy integral formula for the analytic function h from M_1 to E,

$$h(t) = \int_B Q(t,s)h(s)\,dw(s).$$

Since $\{h(s), s \in B\}$ is a bounded set in E_1, the right-hand side of the formula defines an analytic function from C into E_1, and hence h is analytic into E_1.

Proof of Theorem (4.2). We apply the argument of the preceding proof extending T_t^{-1} and f_t to be holomorphic on a complex neighborhood of a point t_0 in the original real-analytic manifold M_1. For each t in this neighborhood, $T_t^{-1}f_t$ is an element of $\mathcal{A}(B)$ since A_t can be extended to the neighborhood to be an elliptic differential operator of order r for each t with coefficients analytic in (x, t). The function $T_t^{-1}f_t$ is an holomorphic function from the neighborhood M'_1 to $L^2(B)$ with values in $\mathcal{A}(B)$. Hence it is a holomorphic function of t with values in $\mathcal{A}(B)$. The analyticity of u over $M \times M_1$ follows from results on the structure of the space $\mathcal{A}(B)$, (cf. [9], p. 315).

5. Elliptic operators depending upon a parameter. We now state and prove the results described in the Introduction conecrning an analytic family of elliptic operators in a vector bundle over a non-compact manifold.

THEOREM 5.1. *Let M and M_1 be real analytic manifolds with M non-compact. Let M' be an open subset of M having compact closure in M, B an analytic vector bundle over M. Suppose that $\{A_t\}$ is an analytic family of analytic elliptic differential operators in B with t running over M_1. Let B' be the restriction of B over M'. Then:*

(a) *There exists a family of solvable realizations $\{T_t\}$ of A_t in $L^2(B')$ such that $(T_t)^{-1}$ is analytic in t on M_1 and has a kernel e_t which is a biregular fundamental solution for A_t over M'. If for a fixed parameter value t_0 in*

M_1, a kernel e_0 is given corresponding to the inverse of a solvable realization T_0 of A_{t_0} in $L^2(B')$, then the family $\{T_t\}$ may be chosen so that $T_{t_0}=T_0$, i. e. $e_{t_0}=e_0$. The kernels e_t are analytic jointly in all variables except over the diagonal in $M'\times M'$. Each kernel e_t lies in $L^1{}_{\mathrm{loc}}(B'\otimes B')$, and the mapping $t\rightarrow e_t$ is an analytic mapping of M_1 into the latter space.

(b) *Suppose that there is given an analytic mapping f of M_1 into $L^2(B')$ such that for every t in M_1, f_t lies in $\mathfrak{A}(B')$. Then there exists an analytic mapping $u\colon M_1\rightarrow L^2(B')$ such that for all t in M_1, u_t lies in $\mathfrak{A}(B')$ and $A_tu_t=f_t$. If for a fixed parameter value t_0 in M_1, we are given a solution u_0 in $L^2(B')$ of the equation $A_{t_0}u_0=f(t_0)$, then the analytic family $u(t)$ can be chosen so that $u(t_0)=u_0$. The family $u(t)$ is analytic simultaneously in all variables.*

Proof of Theorem (5.1). The proof is based upon the application of the results of Section 1 combined with the information already given on elliptic differential operators. Let $A_{t,0}$ and $A_{t,1}$ be the maximal and minimal operators respectively, of A_t for each t in M_1, and let $A'_{t,0}$ be the minimal operator of A'_t, the formal adjoint of A_t with respect to a fixed analytic inner product in B'. By Theorems (4.1) and (4.2), it suffices for the proof of Theorem (5.1) to show the existence of a family of realizations $\{T_t\}$ of $\{A_t\}$ in $L^2(B')$ such that $R_t=(T_t)^{-1}$ varies analytically with t in M_1 as an element of $\mathfrak{L}(L^2(B'),L^2(B'))$. In addition, we must show that if e_0 or u_0 are given as in parts (a) and (b) of Theorem (5.1), that we can construct the family of realizations T_t so the kernel of $(T_{t_0})^{-1}=e_0$, or that $T_{t_0}u_0=f(t_0)$, (i. e. u_0 lies in the domain of T_{t_0}).

To construct this family of realizations, we appeal to Theorems (1.1) and (1.2). In the application of these theorems, we let $H=L^2(B')$ with the given analytic inner product and let $T_{0,t}=A_{t,0}$, $T'_{0,t}=A'_{t,0}$. By Lemma (2.3), the domains of all the operators $T_{0,t}$ and $T'_{0,t}$ thus defined are the same set, $W=W^{r,2}{}_0(B')$. By the definition of the formal adjoint differential operator, $T_{0,t}$ is formally adjoint to $T'_{0,t}$ in the operator sense of Section 1. From the analyticity of the coefficients over M in which M' has compact closure and from the fact that all differential operators of order $\leqq r$ are bounded maps of $W^{r,2}{}_0(B')$ into $L^2(B')$, it follows immediately that $T_{0,t}$ and $T'_{0,t}$ as maps of W into H vary analytically with t in M_1. Finally, by Lemma (2.3) since elliptic operators with analytic coefficients satisfy the condition $(U)_s$ for uniqueness of the Cauchy problem in the small, we know that for each t in M_1, the nullspaces of $T_{0,t}$ and $T'_{0,t}$ are trivial. Hence by Theorem (1.1), there exists a family of realizations $\{T_t\}$ with $A_{t,0}\subseteq T_t\subseteq(A'_{t,0})^*$

$= A_{t,1}$ with T_t^{-1} analytic in t and satisfying the conditions of Theorem (5.1). Hence the proof of Theorem (5.1) is complete.

THEOREM (5.2). *Let M and M_1 be real analytic manifolds with M_1 separable and M non-compact. Let M' be an open subset of M having compact closure in M, M'' an open subset of M' such that* $\mathrm{cl}(M'') \subset M'$ *and such that $M' - M''$ has no compact components. Let B be an analytic vector bundle over M, $\{A_t\}$ an analytic family of analytic elliptic differential operators in B indexed over M_1. Let $h\colon M_1 \to \mathcal{A}(B'')$ be an analytic mapping of M_1 into the analytic functions over B'', where B'' is the restriction of B to M'', such that for each t in M_1, $A_t h(t) = 0$ over M''. Then for each* $\epsilon > 0$, each *integer $j \geqq 0$, each compact subset K'' of M'' and each compact subset K of M_1, there exists an analytic mapping $h'\colon M_1 \to \mathcal{A}(B')$, where B' is the restriction of B over M', such that $A_t h'(t) = 0$ over M', and for $t \in K$ and any normalized differential operator D of order $\leqq j$, we have*

$$|Dh(t) - Dh'(t)| < \epsilon, \text{ on } K''. \tag{5.1}$$

For each fixed value of t, we can choose an element $h'(t)$ of $\mathcal{A}(B')$ which satisfies the inequality (5.1) by Proposition (2.1). It remains to be shown that $h'(t)$ can be chosen so that it too is analytic in t over M_1.

To establish this fact, we note first of all the following side consequence of the argument by which we proved Theorem (5.1).

THEOREM (5.3). *Consider M, M_1, M', B, and the analytic family of elliptic differential operators $\{A_t\}$ as in Theorem* (5.1). *For each t, let S_t be the nullspace of the maximal operator $A_{t,1}$ of A_t in $L^2(B')$. Then the orthogonal projection P_t of $L^2(B')$ onto the closed subspace S_t varies analytically with t in M_1.*

Proof of Theorem (5.3). We make the same identifications in the mechanisms of Theorem (1.1) as was done in the proof of Theorem (5.1). By part (I)(b) of the proof of Theorem (1.1), the orthogonal projection E'_t of $L^2(B')$ onto the range of $A'_{t,0} = T'_{0,t}$ varies analytically with t. However, $P_t = I - E'_t$, as shown in part (I)(c) of that proof, and the theorem follows.

Proof of Theorem (5.2). We begin by showing that it is possible to find $h'(t)$ which is differentiable in t (i.e. infinitely differentiable) and satisfies the other conditions of Theorem (5.2). To simplify the argument, let us note that it suffices to establish the single inequality

$$\| h(t) - h'(t) \|_{L^2(B'')} < \epsilon \tag{5.1$'$}$$

in place of the original inequality (5.1), by a standard argument involving the fact that over the spaces S_t the topologies of $L^2{}_{loc}(B'')$ and $\mathcal{E}(B'')$ are equivalent and indeed uniformly over the compact sets of M_1. Here, without any loss of generality, we have assumed that $h(t)$ for each t actually lies in $L^2(B'')$, as would follow in any case by shrinking M'' slightly so that it still contains K'' and still has the property that $M'-M''$ has no compact components. We shall make this assumption through the remainder of the proof and construct approximations $h'(t)$ which lie in $L^2(B')$.

For each point t in M_1, we may find $k(t)$ by Proposition (2.1) such that $k(t)$ lies in $L^2(B')$, $A_t k(t)=0$ over M', and

$$(5.3) \qquad \| h(t)-k(t)\|_{L^2(B'')} < \epsilon/3.$$

(Note that we can expand M' slightly without changing the hypotheses and obtain $k(t)$ not merely in $L^2{}_{loc}(B')$ as in Proposition (2.1) but actually analytic on the closure of M'.)

Since the orthogonal projection P_t of $L^2(B')$ onto S_t varies analytically with t on M_1, it is continuous in norm by Proposition 1 of Section 1. Hence for each point t in K, there exists a neighborhood $U(t)$ of t in M_1 such that for t' in $U(t)$,

$$(5.3) \qquad \| P_t - P_{t'} \| < \frac{\epsilon}{3\,\| k(t)\|_{L^2(B')}}.$$

Since K is compact, it can be covered by a finite number of these neighborhoods $U_j = U(t_j)$. From (5.3), it follows that for any t in U_j,

$$(5.4) \qquad \| P_{t_j}(k(t_j)) - P_t(k(t_j))\|_{L^2(B')} < \epsilon/3$$

and a fortiori, since $P_{t_j}(k(t_j)) = k_{t_j}$,

$$(5.5) \qquad \| k(t_j) - P_t(k(t_j))\|_{L^2(B'')} < \epsilon/3.$$

Under our assumptions, $h(t)$ can be taken as an analytic mapping of M_1 into $L^2(B'')$, and hence by refinement of the covering, we may assume that for all t in U_j and each j,

$$(5.6) \qquad \| h(t) - h(t_j)\|_{L^2(B'')} < \epsilon/3.$$

We now choose an infinitely differentiable partition of unity $\{\xi_j\}$ corresponding to the covering $\{U_j\}$ of K, with ξ_j differentiable on M_1 and $0 \leqq \xi_j(t) \leqq 1$ for all t in M_1. We define

$$(5.7) \qquad h'(t) = \sum_j \xi_j(t) P_t(k(t_j)).$$

The sum is finite and obviously $h'(t)$ is infinitely differentiable in t in M_1 and lies in S_t for each t, since each of the summands does. To complete this part of the argument, we need to show that $h'(t)$ satisfies the inequality (5.1)$'$. However,

$$h(t)-h'(t)=\sum_j \xi_j(t)\{h(t)-P_t(k(t_j))\},$$

while

$$h(t)-P_t(k(t_j))=\{h(t)-h(t_j)\}+\{h(t_j)-k(t_j)\}+\{k(t_j)-P_t(k(t_j))\}. \tag{5.8}$$

For t in the support of ξ_j which is contained in U_j, each of the three expressions on the right of equation (5.8) has norm in $L^2(B'')$ less than $\epsilon/3$ by (5.2), (5.5), and (5.6). Hence

$$\| h(t)-P_t(k(t_j))\|_{L^2(B'')}<\epsilon$$

for all t in U_j. Thus $h'(t)-h(t)$ is a convex linear combination of elements in the open ϵ-ball about the origin $L^2(B'')$, so that

$$\| h'(t)-h(t)\|_{L^2(B'')}<\epsilon \tag{5.9}$$

as we wished to show.

To complete the prof of Theorem (5.2), it suffices to approximate the family $h'(t)$ just constructed by an analytic family $h''(t)$ so that

$$\| h'(t)-h''(t)\|_{L^2(B')}<\epsilon. \quad t\in K, \tag{5.10}$$

with the approximating $h''(t)$ also lying in S_t for each t in M_1. Let us note however, that the restriction that $h''(t)$ should lie in S_t is actually inessential since if we had an analytic family $h''(t)$ in $L^2(B')$ satisfying (5.10), then setting $h_1(t)=P_th''(t)$, we should have for t in K,

$$\begin{aligned}\| h'(t)-h_1(t)\|_{L^2(B')}&=\| P_t(h'(t)-h''(t)\|_{L^2(B')}\\ &\leqq \| h'(t)-h''(t)\|_{L^2(B')}<\epsilon.\end{aligned}$$

Thus it suffices to approximate h' by an analytic mapping of M_1 into $L^2(B')$. This follows easily from the Morrey-Grauert Imbedding Theorem by imbedding M_1 analytically as a closed submanifold of an Euclidean space E^n, considering $h'(t)$ as a continuous function with compact support from M_1 to $L^2(B')$, extending h' to a continuous function with compact support from E^n to $L^2(B')$ by a simple variant of Tietze's theorem, and finally taking the convolution of the extended function with an analytic function of the form

$$c\delta^{-n}\exp(-|x|^2/\delta^2)$$

(with c adjusted to give L^1-norm equal to 1) for δ sufficiently small. The details of the argument are standard and will be omitted.

6. Non-elliptic operators. With suitable modifications, the results which are obtained for elliptic operators in Section 5 can be extended in large part to some significant classes of non-elliptic operators. We shall describe this extension in its most abstract form and then specialize to those classes of differential operators for which our abstract hypotheses have been shown to hold.

Let B be an analytic vector bundle over the manifold M, M_1 a real analytic manifold, $\{A_t\}$ as before an analytic family of differential operators on M indexed by t in M_1 (and not necessarily elliptic). Let A'_t be the formal adjoint of A_t with respect to a fixed analytic inner product in B, $L^2(B)$ the corresponding L^2 space with respect to this inner product.

DEFINITION 6.1. (a) *The minimal operator $A_{t,0}$ of A_t in $L^2(B)$ is the closure as an operator in $L^2(B)$ of A_t restricted to $\mathfrak{D}(B)$, $(t \in M_1)$. Similarly $A'_{t,0}$ is the minimal operator of A'_t in $L^2(B)$.*

(b) *The maximal operator $A_{t,1}$ of A_t in $L^2(B)$ is defined by $A_{t,1} = (A_{t,0})^*$.*

(c) *The closed operator T in $L^2(B)$ is said to be a solvable realization of A_t for a fixed t in M_1 if $A_{t,0} \subseteq T \subseteq A_{t,1}$ while T has a bounded inverse R defined on all of $L^2(G)$.*

DEFINITION 6.2. *The differential operator A_t is said to be proper if there exists a constant $c(t)$ such that*

$$\| u \| \leqq c(t) \| A_t u \|$$

for all u in $\mathfrak{D}(B)$. (All norms except where otherwise indicated are taken in $L^2(B)$.)

Equivalently, A_t is proper if $A_{t,0}$ has a trivial null space and a closed range.

DEFINITION 6.3. *The family $\{A_t\}$ is said to be analytically uniform in $L^2(B)$ if there exists a fixed linear subset W of $L^2(B)$ and a function U from M_1 to the bounded linear operators with bounded inverses on $L^2(B)$ such that for each fixed t in M_1, U_t maps W onto the domain of $A_{t,0}$ while for each fixed w in W, $A_{t,0}U_t w$ is an analytic function from M_1 to $L^2(B)$.*

The most important special case of Definition (6.3) is that in which $W = D(A_{t,0})$, where the latter domain of $A_{t,0}$ is independent of t in M_1.

THEOREM 6.1. *Let $\{A_t\}$ be a family of differential operators on B*

depending upon t in M_1. Suppose that each A_t and A'_t is proper in $L^2(B)$ and that the families $\{A_t\}$ and $\{A'_t\}$ are analytically uniform. Then:

(a) *There exists a family of closed linear operators $\{T_t, t \in M_1\}$ in $L^2(B)$ such that each T_t is a solvable realization of A_t in $L^2(B)$ while $(T_t)^{-1}$ is analytic in t on M_1. If t_0 is a given point of M_1 and T_0 a given solvable realization of A_{t_0} in $L^2(B)$, the family $\{T_t\}$ may be chosen so that $T_{t_0} = T_0$.*

(b) *The kernels e_t of the mappings $(T_t)^{-1}$ are an analytic family of distribution fundamental solutions for A_t on B,*

$$A'_t, \bar{e}_t(\cdot, y) = \delta_y,$$
$$A_t e_t(x, \cdot) = \delta_x.$$

(c) *If f is an analytic map from M_1 into $L^2(B)$, then $u_t = (T_t)^{-1} f_t$ gives an analytic family of solutions of the equations $A_t u_t = f_t$ in $L^2(B)$. If u_0 is a given solution of $A_{t_0} u_0 = f_{t_0}$ for a given value t_0 in M_1, then the family T_t may be chosen so that $u_{t_0} = u_0$.*

THEOREM 6.2. *Let $\{A_t\}$ be a family of differential operators in B parametrized by t in M_1 such that each A'_t is proper and the family $\{A'_t\}$ is analytically uniform. For each fixed t, let S_t be the closed subspace of $L^2(B)$ consisting of solutions u of the homogeneous equation $A_t u = 0$ (in the distribution sense). Let P_t be the orthogonal projection operator of $L^2(B)$ on S_t. Then the mapping $t \to P_t$ is analytic from M_1 into the projections of $L^2(B)$. In particular, P_t is continuous in t in the norm topology so that the angle between S_t and S_{t_0} approaches zero as $t \to t_0$.*

The proofs of Theorems (6.1) and (6.2) are simple applications of a slight variant of the arguments used to prove Theorems (1.1) and (1.2) of Section 1. We shall prove Theorem (6.2) first and then apply it to the proof of Theorem (6.1).

Proof of Theorem (6.2). Since A'_t is proper for each t in M_1, the minimal operator $A'_{t,0}$ has closed range R'_t in $L^2(B)$ and trivial nullspace. If E'_t is the orthogonal projection on R'_t, it follows by an easy argument that $P_t = (I - E'_t)$. Hence it suffices to prove E'_t analytic in t.

Let U_t be the operators in the analytic uniformity assumption on A'_t. Let t_0 be a given point of M_1, and let $R'_0 = R'_{t_0}$. For each t in M_1, we define the linear mapping K_t of R'_0 into $L^2(B)$ with range R'_t by setting $K_t = A'_{t,0} U_t (U_{t_0})^{-1} (A'_{t_0,0})^{-1}$. For fixed t in M_1, K_t is a closed linear operator from R'_0 to $L^2(B)$. Indeed, let $u_m \to u$, $K_t u_m \to v$. Since by the properness of A'_t, $A'^{-1}_{t_0}$ is continuous from R'_0 to $L^2(B)$,, we know that

$$(A'_{t_0,0})^{-1} u_m \to (A'_{t_0,0})^{-1} u.$$

U_t and $(U_{t_0})^{-1}$ are both continuous by assumption. Hence

$$U_t(U_{t_0})^{-1}(A'_{t_0,0})^{-1}u_m \rightarrow U_t(U_{t_0})^{-1}(A_{t_0,0})^{-1}u.$$

Finally $A'_{t_0,0}$ is a closed operator, so that

$$K_t u_m = A'_t U_t (U_{t_0})^{-1}(A'_{t_0,0})^{-1}u_m \rightarrow v$$

implies that

$$v = A'_t U_t (U_{t_0})^{-1}(A'_{t_0,0})^{-1}u = K_t u.$$

Thus K_t is closed.

By the closed graph theorem, K_t is a bounded linear operator. For fixed u in R'_0 and variable t, we know (setting $w = (U_{t_0})^{-1}(A'_{t_0,0})^{-1}u$) that $K_t u = A'_{t_0} U_t w$ is analytic in t because of the analytic uniformity assumption upon the family $\{A'_t\}$. Hence K_t is weakly analytic and by Proposition 1 of Section 1, K_t is analytic in the norm topology. In particular, K_t is continuous in the norm topology. For $t = t_0$, K_{t_0} reduces to the identity mapping of R'_0 into $L^2(B)$. Hence $K_t = I + Q_t$, where Q_t is analytic in t and $Q_{t_0} = 0$. Applying Lemma 2 of Section 1, it follows that the orthogonal projection operators E'_t of $L^2(B)$ on R'_t are analytic near t_0. Since t_0 was arbitrary in M_1, the proof of Theorem (6.2) is complete.

COROLLARY TO THEOREM 6.2. *Under the hypotheses of Theorem* (6.1), *the orthogonal projections* E_t, E'_t, P_t, *and* P'_t *of* $L^2(B)$ *on* $R(A_{t,0})$, $R(A'_{t,0})$, $N(A_{t,1}($, *and* $N(A'_{t,1})$, *respectively, are all analytic in* t.

Proof of the Corollary. In the proof of Theorem (6.2), we showed that if each A'_t is proper and if $\{A'_t\}$ is analytically uniform, then E'_t and P_t are analytic in t on M_1. Reversing the roles of A_t and A'_t, it follows that of each A_t is proper and the family $\{A_t\}$ is analytically uniform, then E_t and P'_t are analytic in t on M_1. Under the hypotheses of Theorem (6.1), both hypotheses hold so that all four operators are analytic in t.

Proof of Theorem (6.1). Let $Y_t = (A'_{t,0})^{-1}E'_t$. By the argument of part (II) of the proof of Theorem (1.1), Y_t is analytic in t on M_1, and $A_{t,1}Y_t^* = I$. We construct the solvable realization T_t which we seek by setting

$$T_t = (A_{t,0})^{-1}E_t + Y_t^*(I - E_t).$$

The modification of the family $\{T_t\}$ to pass through a given realization T_0 or to have the property that $T_{t_0}u_0 = f_{t_0}$ is carried through as in the proofs of Theorems (1.1) and (1.2).

Let us examine briefly some significant classes of differential operators for which the assumptions of Theorems (6.1) and (6.2) can be verified. We assume that M is an open subset G of E^m where G has compact closure contained in an open set G_e such that the coefficients a_α and a'_α can be extended to be analytic on $G_e \times M_1$. (This assumption is un-necessarily strong but avoids any complication in detailing differentiability hypotheses). Then (considering only scalar operators):

(1) Every operator with constant coefficients is proper [11];

(2) Every elliptic operator is proper. The same holds true for domains G satisfying suitable 'convexity' conditions for hyperbolic operators (Leray [15]), differential operators not having multiple complex characteristic roots in a given direction (Calderon [6]), or the slightly larger class of differential operators of principal normal type having a strongly pseudo-convex level function on G_e (Hörmander [11]). In all these cases, the proper character of A_t follows from the fact that the following two conditions are verified:

(i) $D(A_{t,0})$ is contained in a subset W of $L^2(G)$ which is a Banach space with respect to a norm such that the injection map into $L^2(G)$ is a compact map.

(ii) Every solution u in $L^2(G)$ of the equation $A_t u = 0$ on G^e such that u has compact support in G^e must be null.

For the subset W, we may take $W^{j,2}{}_0(G)$ with $j > 0$, and the corresponding results about domains of minimal operators and about uniqueness of the Cauchy problem follows for suitable domains G and G^e.

(3) Some examples of families $\{A_t\}$ which are analytically uniform by having a fixed domain for the minimal operators $A_{t,0}$ are the following: elliptic operators of a fixed order; formally hypoelliptic operators of equal strength for different values of x and t; constant coefficient operators of equal strength; hyperbolic operators (and more generally operators of the type studied by Calderon and Hörmander in the references cited above) with top order terms independent of t.

We close with some final remarks: (1) J. F. Treves [20] has recently studied from another point of view, and using completely different methods, constant coefficient operators varying analytically with t and has obtained in the case of operators of equal strength elementary solutions analytic in t near a fixed value t_0; (2) The results of this Section may be extended with purely formal changes to obtain holomorphic dependence for the parameter running over a complex manifold.

YALE UNIVERSITY.

REFERENCES.

[1] F. E. Browder, "Functional analysis and partial differential equations, I," *Mathematische Annalen*, vol. 138 (1959), pp. 55-79.

[2] ———, "On the spectral theory of elliptic differential operators, I," *Mathematische Annalen*, vol. 142 (1961), pp. 22-130.

[3] ———, "Functional analysis and partial differential equations, II," *Mathematische Annalen*, vol. 145 (1962), pp. 81-226.

[4] ———, "Functional analysis and partial differential equations, III," *to appear.*

[5] ———, "On analyticity and partial differential equations," *Bulletin of the American Mathematical Society*, vol. 68 (1962), pp. ———.

[6] A. P. Calderon, "Existence and uniqueness theorems for systems of partial differential equations," *to appear.*

[7] N. Dunford and J. T. Schwartz, *Linear Operators, I*, New York, 1958.

[8] I. M. Gelfand, "Abstrakte Funktionen und lineare Operatoren," *Matematičeskiĭ Sbornik Novaya Serie*: (Moscow) vol. 4 (1938), pp. 235-286.

[9] A. Grothendieck, *Espaces Vectoriels Topologiques*, San Paulo, 1954.

[10] E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, American Mathematical Society Colloquium Publications, vol. 31, Providence, 1957.

[11] L. Hörmander, "On the theory of general partial differential operators," *Acta Mathematica*, vol. 94 (1955), pp. 161-248.

[12] L. Hörmander, "Differential operators with non-singular characteristics," *Bulletin of the American Mathematical Society*, vol. 68 (1962), pp. 354-359.

[13] K. Kodaira and D. C. Spencer, "On deformations of complex structures III, Stability theorems for complex structures," *Annals of Mathematics*, vol. 71 (1960), pp. 43-76.

[14] P. Lax, "A stability theory of abstract differential equations and its application to the study of local behaviour of solutions of elliptic equations," *Communications on Pure and Applied Mathematics*, vol. 8 (1956), pp. 747-766.

[15] J. Leray, *Hyperbolic equations*, Institute for Advanced Study Lecture Notes, Princeton, 1953.

[16] J. Lions and L. Schwartz, "Problèmes aux limites sur certains espaces fibrés," *Acta Mathematica*, vol. 94 (1955), pp. 155-159.

[17] B. Malgrange, "Existence et approximation des solutions des équations aux derivées partielles et des équations de convolution." *Université Grenoble, Annales de l'Institut Fourier*, vol. 6 (1956), pp. 271-355.

[18] L. Schwartz, *Théorie des distributions*, Paris 1950-1, 2 vols.

[19] ———, "Théorie de noyaux," *Proceedings of the International Congress of Mathematicians*, Cambridge, 1950, vol. 1, pp. 220-230.

[20] J. F. Treves, "Fundamental solutions of linear partial differential equations with constant coefficients depending on parameters," *to appear.*

[21] M. I. Visik, "On general boundary problems for elliptic differential equations," *Trudy Moskovskogo Matematičeskogo Obščestva (Moscow)*, vol. 1 (1952), pp. 187-246.